PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :		(11) International Publication Number: WO 92/21753
C12N 15/11, 9/50, 15/29 C12N 1/21, A01H 1/06	A1	(43) International Publicati n Date: 10 December 1992 (10.12.92)

(21) International Application Number: PCT/US92/04785

(22) International Filing Date: 8 June 1992 (08.06.92)

(30) Priority data:
712,024 7 June 1991 (07.06.91) US

(71) Applicant: DOWELANCO [US/US]; 9002 Purdue Road, Indianapolis, IN 46268-1189 (US).

(72) Inventors: WALSH, Terence, A. ; 2335 North Trail, Midland, MI 48642 (US). OWENS MERLO, Patricia, Ann ; 4612 Spring Brook Court, Midland, MI 48642-6808 (US). STRICKLAND, James, A. ; 4314 Paine Drive, Midland, MI 48642-5820 (US). ORR, Gregory, L. ; 206 Wallace Street, P.O. Box 39, Ashley, MI 48806-0039 (US). MERLO, Donald, J. ; 4612 Spring Brook Court, Midland, MI 48642-6808 (US). WALDRON, Clive ; 6206 Green Leaves Road, Indianapolis, IN 46220-4811 (US).

(74) Agent: ZINDRICK, Thomas, D.; DowElanco, Patent Department, P.O. Box 1266, Midland, MI 48641-1266 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BG, BR, CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LK, LU (European patent), MC (European patent), MG, MW, NL (European patent), NO, PL, RO, RU, SD, SE (European patent).

Published
With international search report.

(54) Title: INSECTICIDAL PROTEINS AND METHOD FOR PLANT PROTECTION

(57) Abstract

The present invention provides a composition and method of using certain cysteine protease inhibitors to protect plants otherwise susceptible to insect infestation by one or more of Mexican bean beetle, red flour beetle, confused flour beetle, cowpea beetle, boll weevil, Colorado potato beetle, three-lined potato beetle, rice weevil, maize weevil, granary weevil, lesser grain borer, flea beetles, Egyptian alfalfa weevil, bean weevil, yellow mealworm, asparagus beetle and squash bug.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FI	Finland	ML	Mali
AU	Australia	FR	France	MN	Mongolia
BB	Barbados	GA	Gabon	MR	Mauritania
BE	Belgium	GB	United Kingdom	MW	Malawi
BF	Burkina Faso	GN	Guinea	NL	Netherlands
BG	Bulgaria	GR	Greece	NO	Norway
BJ	Benin	HU	Hungary	PL	Poland
BR	Brazil	IE	Ireland	RO	Romania
CA	Canada	IT	Italy	RU	Russian Federation
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland			SN	Senegal
CI	Côte d'Ivoire	KR	Republic of Korea	SU	Soviet Union
CM	Cameroon	LI	Liechtenstein	TD	Chad
CS	Czechoslovakia	LK	Sri Lanka	TG	Togo
DE	Germany	LU	Luxembourg	US	United States of America
DK	Denmark	MC	Monaco		
ES	Spain	MG	Madagascar		

INSECTICIDAL PROTEINS AND METHOD FOR PLANT PROTECTION

The present invention relates to the fields of genetic engineering and plant husbandry. More specifically, the invention provides methods and compounds for controlling or combating insects in agriculture or horticulture.

Many vegetable and field crops are attacked by insect pests. Most plants show some resistance to certain insects; the resistance can be physical or chemical. For example, the hairs on the leaves of many plants can stop small insects from getting near enough to the surface to chew it. In other cases plants use a range of complex secondary chemicals to make their tissues unattractive or toxic. Control ofsuchphytophagous insects has traditionally been partially addressed by cultural and breeding methods. An effective way to reduce these losses is to use crop cultivars having genes for pest resistance (see Painter (1951), Insect Resistance in Crop Plants, Macmillan: New York). Plant breeders have attempted to reduce losses caused by insect attack by incorporating insect resistance genes into their varieties via conventional breeding programs.

Classical approaches to host plant resistance, though remarkably successful in some instances, are rather empirical. Once "traits" for resistance are discovered, they are moved into agronomically acceptable lines by selection procedures. One limitation of the classical approach is that the movement of genes for resistance from one plant to another is restricted to species that can be interbred. Additionally, these types of resistance are likely to be under the control of many genes, and so are difficult for the plant breeder to fully exploit. Often resistant varieties have shown a yield depression and so have not been economically viable. Moreover, if no resistance can be identified within a species or within related species, then no improvement in insect pest resistance is possible by classical breeding.

Chemical insecticides have been heavily relied upon to control insects. These agents typically are applied on or banded into the soil, or to plant foliage or in bait stations. In spite of the availability of a wide range of chemical pesticides, phytophagous insects remain a serious problem. Many chemical pesticides have the disadvantage of requiring repeated applications. A major problem in the use of many pesticides is the ability of insects to become resistant to the applied agents. This phenomenon occurs through selection of the most resistant members of the insect population during repeated application of the agent. A need therefore exists for new insect control agents, particularly agents that have a mode of action different from conventional insecticides.

As alternatives to synthetic compounds, certain naturally-occurring agents have been isolated and

developed as pesticides. These include plant and microbial secondary metabolities and proteins, and natural predators or pathogens of insects (including other insects, fungi, bacteria and viruses). Furthermore, as recombinant DNA technology has advanced, genes from a donor organism may be transferred to a recipient organism resulting in a new phenotype in the recipient. In the case of transgenic plants, this phenotype may be resistance to insect damage if the introduced gene encodes a polypeptide, the action of which results in a deleterious effect on the pest. Consequently, there is a great interest and utility in finding polypeptides that have such an effect. Genes for these polypeptides can be used to modify organisms, especially plants and microbess, so that they adversely affect the growth and development of insect pests. A very limited number of such polypeptides have been described, e.g., polypeptides from *Bacillus thuringiensis* various proteinaceous protease and amylase inhibitors, various plant lectins, etc.

One physiological system of insects known to be susceptible to disruption by specific inhibitors is the action of digestive proteases. The digestive proteases hydrolyze ingested proteins and polypeptides by cleaving peptide bonds. The term "protease" is specifically intended to include endopeptidases and exopeptidases of the four major catalytic classes: serine proteases, cysteine proteases, carboxyl proteases and metallo proteases (see Laskowski *et al.* (1983), Ann. Rev. Biochem., 49:593-626). The class to which a specific protease belongs can be determined by the pH range over which it is active, by its ability to hydrolyze specific

proteins, by its similarity to other well-characterized proteases and by its sensitivity to various inhibitors.

Diverse types of insect digestive enzymes release peptides and amino acids from dietary protein. One class of digestive enzymes is the cysteine proteinases. The term "cysteine proteinase" is intended to describe a protease that possesses a highly reactive thiol group of a cysteine residue at the catalytic site of the enzyme. There is evidence that many phytophagous insects rely, at least in part, on midgut cysteine proteases for protein digestion. These include *Hemiptera*, especially squash bugs (*Anasa tristis*); *Coleoptera*, especially, corn rootworm (*Diabrotica* spp.); Mexican bean beetle (*Epilachna varivestis*); red flour beetle (*Tribolium castaneum*); confused flour beetle (*Tribolium confusum*); cowpea weevil (*Callosobruchus maculatus*); boll weevil (*Anthonomus grandis*); Colorado potato beetle (*Leptinotarsa decemlineate*); three-lined potato beetle (*Lema trilineata*); rice weevil (*Sitophilus oryzae*); maize weevil (*Sitophilus zeamais*); granary weevil (*Sitophilus granarius*); lesser grain borer (*Rhyzopertha dominica*); the flea beetles (*Chaetocnema spp.*, *Haltica spp.*, and *Epitrix spp.*); Egyptian alfalfa weevil (*Hypera postica*); bean weevil (*Acanthoscelides obtectus*); yellow mealworm (*Tenebrio molitor*); and asparagus beetle (*Crioceris asparagi*).

Compounds that form complexes with proteases and inhibit their proteolytic activity are widespread in nature. A variety of "low molecular weight" proteinase inhibitors are known, largely of non-natural synthetic origin. A number of naturally occurring low molecular weight inhibitors have been isolated from bacterial and fungal sources and characterized; this group includes such inhibitors as E-64 (N-(L-3-trans-carboxyoxiran-22-

carbomyl)-L-leucyl)-amido-4-guanido) butane), the leupeptins, antipains and pepstatins.

Several proteinaceous proteinase inhibitors have been isolated from plant species and are among the defensive chemicals in plant tissues that are both developmentally regulated and induced in response to insect and pathogen attacks. Inhibitors of serine-, cysteine-, aspartic acid-, and metallo-proteinases have been found in plants and especially in storage organs such as tubers and seeds. The most common and widely studied group of plant protease inhibitors are those that inhibit the animal serine proteases, which include trypsin and chymotrypsin (see Ryan (1990), Annu. Rev. Phytopatnol, 28:425-449).

Proteinaceous cysteine protease inhibitors decrease or eliminate the catalytic activity of a cysteine protease. The pH optima of cysteine proteinases is usually in the range of 5-7, which is the pH range in the lumen of midguts of insects that use cysteine proteinases (see Ryan, (1990), *supra*). Cystatins are naturally occurring, proteinaceous cysteine proteinase inhibitors (see Barrett (1987), Trends. Biol. Sci., 12:193-196). Cystatins have been classified into three families with respect to molecular weight, the number of disulfide bonds, subcellular localization, and primary structure characteristics (see Barrett (1987), *supra*). The classification system is mainly based on information regarding vertebrate cystatins.

One family comprises the Type 1 cystatins (sometimes called stefins) whose molecules consist of a single domain of about 100 amino acid residues (M_r

11,000) with no disulfide bonds or carbohydrate groups. Somewhat more complex are the Type 2 cystatins, which are molecules having a single domain of about 115 amino acid residues (M_r 13,000), and contain two disulfide loops near the carboxyl terminus. Chicken cystatin is of this type, as are human cystatins C and S, a cystatin from beef colostrum, and à cystatin from snake venom. The most complex cystatin molecules are those of the Type 3 cystatins, the kininogens. Each of these contains four domains: domain 4 is a kinin (which is not inhibitory of cysteine proteases) and domains 1-3 are cystatin-like domains that have apparently resulted from two duplications of genetic material; however, domain 1 is not inhibitory to cysteine proteases, whereas domains 2 and 3 are inhibitory.

The regions of cystatins responsible for inhibitory activity can be summarized as follows: sequence homology among members of the cystatin superfamily reveals two well conserved features; one, the amino-terminal Gly residue and the other, the sequence Gln-Val-Val-Ala-Gly or its analogue Gln-Xxx-Val-Xxx-Gly. This highly conserved region is considered to construct the "binding edge" with other conserved segments and to interact with cysteine proteinase at the active site cleft.

It has been taught that animal cystatins (such as hen egg white cystatin and kininogens) and low molecular weight, nonpeptide cysteine protease inhibitors (such as E-64, antipain and leupeptin) may be effective in the control of a variety of *Coleoptera* which utilize cysteine proteinases (see Australian Patent Application No. 36568/89).

Relatively little is known about plant cysteine protease inhibitors. Cysteine proteinase inhibitors have been reported from a variety of plant tissue, including pineapple (see Reddy *et al.* (1975), J. Biol. Chem., 250:1741-1750); potato, also commonly referred to as potato papain inhibitor (PPI) (see Rodis and Hoff (1984), Plant Physiol., 74:907-911); corn (see Abe *et al.* (1988), Agric. Biol. Chem., 52:1583-15843); rice (see Abe *et al.* (1987), J. Biol. Chem., 262:16793-16797); cowpea (see Rele *et al.* (1980), Arch. Biochem. Biophys., 204:117-128); mung bean (see Baumgartner *et al.* (1976), Plant Physiol., 58:1-6); tomato (see Akers *et al.* (1980), Can. J. Bot., 58:1000-1003), wheat, barley and rye (see Fossom (1970), Acta Pathol. Microbial. Scand. Sec B Microbial., 78:741-754); millet (see Tashiro and Maki (1986), Agric. Biol. Chem., 50:2955-2957); pumpkin (see Valveski *et al.* (1991), Plant Science, 74:179-184); Scots pine (see Salmia (1980), Physiol. Plant., 48:266-270); and *Enterolobium contortisiliquum* beans (see Oliva *et al.* (1988), Biol. Chem. Hoppe-Seyler, 369:229-232).

Oryzacystatins I and II from rice seeds have been identified as cystatins. Alignment of the amino acid sequence with those of other cystatins reveals that oryzacystatin is homologous to members of the cystatin superfamilies of animal origin and suggests that this plant cystatin and the animal cystatins have evolved from a cognate ancestral gene (see Abe *et al.* (1988), *supra*).

Partial amino acid sequence data from cysteine proteinase inhibitors purified from seeds of *Wisteria floribunda* (Hirashiki *et al.* (1990), J. Biochem., 108:604-608) and soybeans (Brzin *et al.* (1990), Biol. Chem. Hoppe-Seyler, 371:167-170) also indicate that these are

members of the cystatin superfamily having single cystatin domains.

Protein crystals have long been known to occur naturally in potato tubers. The crystals occur primarily in the subphellogen layer, but sometimes are found throughout the tuber. Structurally intact cubical crystals isolated from the outer cortex of potato tubers are proteins with no significant amounts of nucleic acid, lipid, or carbohydrate (see Hoff *et al.*, (1972), Biochem. Biophys. Res. Commun., 49: 1525-1529). The crystals vary in size from 5 to 25 µm and usually have a cuboidal shape, although sometimes can occur as twin crystals and crystals of prismoidal shape (see Rodis and Hoff (1984), *supra*).

The potato crystalline protein possesses proteinase inhibitor activity against the sulfhydryl proteinases papain, chymopapain, and ficin (see Rodis and Hoff (1984), *supra*). The polypeptide is also unique in that it readily crystallizes at alkaline pH and in the presence of phosphate (see Rodis and Hoff (1984) *supra*), and is produced by potato plants in large quantities in the tuber periderm and leaves.

For example, crystals found in young tomato leaflets, in which chymopapain inhibitor activity was detected, are similar in size and appearance to those observed in potato tubers (see Akers and Hoff (1980), Can. J. Bot., 58:1000-1003). Cytoplasmic crystals also were shown in electron micrographs of tomato leaf tissue after 20 hours of induction.

The function of plant cysteine protease inhibit rs are not well understood in terms of

physiological function. In contrast with trypsin inhibitors, the low and constant levels found for cysteine proteinase inhibitors suggest a participation in metabolic events in the seed. These inhibitors are probably important physiologically and some inhibitors may have nutritional significance (see Valevski *et al.* (1991), Plant Science, 74: 179-184). A protective role against insect and/or pathogens has not been substantiated.

While Ryan lists a number of identified animal and plant proteins that inhibit the cysteine mechanistic class of enzymes, he also states that the vast majority of the inhibitor genes available in nature still remain to be studied and tested for their possible defensive roles in plants (see Ryan (1990), Annu. Rev. Phytopathol., 28:425-49). Thus, there is little direct evidence that the cysteine protease inhibitors, particularly animal cystatins, have a deleterious effect on insect growth and development. In other words, while cysteine protease inhibitors would be expected to inhibit cysteine proteases in an *in vitro* assay of cysteine proteases, there simply is insufficient published data to predict whether a cysteine protease inhibitor would function in a selected insect gut environment to control the growth and development of an insect pest. For example, there is no correlation between the levels of papain inhibitors in selected cowpea seeds and the resistance to *Callosobruchus, maculatus* an insect with predominantly cysteine proteases in the midgut (see Xavier-Filho *et al.* (1989), J. Agric. Food Chem, 37:1139-1143).

In fact, Ryan specifically cites a reference (Murdock *et al.* (1987), Comp. Biochem. Physiol. B, 87:783-

87) which teaches a larval midgut assay of several insects. One *Coleoptera* insect, *T.castaneum*, is shown to have cysteine gut proteases highly inhibited by the alkylating agent pCMB. Indeed, Liang *et al.* (1991) (FEBS, 278(2):139-142) and Chen *et al.* (1992) (Protein Expression and Purification, 3:41-49) have shown that oryzacystatin is a potent inhibitor of *T.castaneum* midgut proteases in an *in vitro* assay. However, Chen *et al.* (1992, (*supra*)) demonstrate that oryzacystatin produces only marginal inhibition of insect growth (35% after 24 days) when oryzacystatin is included in the diet at extremely high levels (10% weight/weight). The effect of 10% oryzacystatin in the diet is considerably less than the effect of non-proteinaceous cysteine protease inhibitor E-64.

An object of the present invention is to provide a method for protecting a plant or a part thereof from insect infestation by insects having digestive cysteine proteases.

A further object of the present invention is to provide novel compositions which are capable of protecting from attack a plant or a part thereof otherwise susceptible to insect infestation by insects having digestive cysteine proteases.

A further object of the present invention is to provide a process for preparing genetically transformed host cells which process comprises the transformation of host cells with a gene encoding a protein capable of causing a deleterious effect, upon ingestion, to insects having digestive cysteine proteases.

Other objects and advantag s of the present invention will become apparent from the description of the invention provided hereunder.

Accordingly in one aspect, the invention relates to a method of controlling insect infestation by insects having digestive cysteine proteases. It is especially concerned with providing amidgut-effective plant cystatin in, on or near plant tissue otherwise susceptible to infestation by such insects, whereby the plant tissue has improved resistance to such insects.

In a second aspect, the present invention relates to an insecticidal composition having a midgut-effective plant cystatin, wherein the composition is capable of improving the resistance of plant tissue otherwise susceptible to insect infestation by one or more insects having digestive cysteine proteases.

In a third aspect, the invention relates to vectors encoding and capable of expressing a midgut-effective plant cystatin in a plant cell.

In a fourth aspect, the invention relates to transformed cells and cell cultures of cells which possess genes encoding a midgut-effective plant cystatin capable of protecting plant tissue otherwise susceptible to insect infestation by one or more insects having digestive cysteine proteases.

In fifth aspect, the present invention relates to a process of preparing an insecticidal composition of a midgut-effective plant cystatin, wherein the composition is capable of improving the resistance of plant tissue otherwise susceptible to insect infestation

by one or more insects having digestive cysteine proteases.

A number of aspects of the present invention are further illustrated in the accompanying Drawings, in which:

Figure 1 shows the deduced amino acid sequence of potato papain inhibitor.

Figure 2 shows the nucleotide sequence of the PPI gene in phage BZ2.

Figure 3 shows the N-terminal amino acid sequences of polypeptides obtained by digestion of PPI with trypsin.

Figure 4 shows a comparison of amino acid sequences of PPI-5, oryzacystatin-1, and hen egg cystatin.

Figure 5 shows effect of PPI on Growth of Second Instar Western and Southern Corn Rootworm.

Figure 6 shows effect of PPI on Growth of Neonate Southern Corn Rootworm.

Figure 7 shows the plasmid map of pDAB219Δ.

Figure 8 shows the plasmid map of pH707-Not.

The entire teachings of all references cited herein are hereby incorporated by reference.

As discussed previously, the mere fact a protease inhibitor is inhibitory in an *in vitro* assay of digestive assays does not necessarily mean it will be effective *in vivo*. This is especially shown in the work

of Purcell *et al.* ((1992), Insect Biochem. Molec. Biol., 22:41-47) as no inhibition of growth of boll weevil or tobacco budworm larvae was associated with feeding high levels of trypsin inhibitors, despite the fact that these proteins were good inhibitors (60-70% inhibition) of midgut proteolytic activity *in vitro*. There is little empirical evidence as to whether specific cystatins have a deleterious effect on selected insect growth and development. What evidence is available suggests previously characterized cystatins are relatively ineffective at controlling selected insect growth and development (see Chen *et al.* (1992), *supra*).

The present invention is directed to an insect controlling amount of midgut-effective plant cystatins with an ability to control infestation by insects having digestive cysteine proteases. Specifically susceptible insects include one or more insects (including larvae) of corn rootworm, Mexican been beetle, red flour beetle, confused flour beetle, cowpea beetle, boll weevil, Colorado potato beetle, three-lined potato beetle, rice weevil, maize weevil, granary weevil, lesser grain borer, flea beetles, Egyptian alfalfa weevil, bean weevil, yellow mealworm, asparagus beetle and squash bug.

An "insect controlling amount" is an amount of a midgut-effective plant cystatin sufficient to deleteriously disrupt the normal life processes of an insect [i.e., amounts which are lethal (toxic) or sublethal (injuring, growth or development inhibiting or repelling)].

As discussed previously, the ability to inhibit proteolytic enzymes *in vitro* does not necessarily mean the

inhibitor will be effective in controlling insect growth. Certain insects having cysteine proteases also possess other digestive proteases which may inactivate ingested cysteine protease inhibitors. While not intended to be bound by theory, it is believed that single domain cystatins are inactivated in the midgut environment prior to being able to bind to and inhibiting the target cysteine proteases of a given target pest.

The present invention is thus directed to a midgut-effective plant cystatin. By "midgut-effective plant cystatin" is meant a cystatin which is resistant to inactivation in the midgut of a target pest for a time sufficient to permit inactivation of digestive cysteine proteases. For purposes of this invention, a "target insect" is one which possesses a cysteine protease as a digestive enzyme. By "resistant" is meant a midgut-effective plant cystatin which is not susceptible to substantial inactivation in the target insect midgut prior to controlling the insect pest.

In a preferred embodiment, the present invention is directed to crystalline cystatins from potato, also referred to as potato papain inhibitor. For purposes of this invention, "potato papain inhibitor" is meant to include a protein encoded by a gene having the sequence set forth in Figure 1, or a functional derivative thereof. As can be seen from Figure 1, the protein component of crystals derived from potato tubers as well as potato leaves is composed of eight domains that, while intending to be bound by theory, appear to have resulted from the duplication of genetic material. The ight domains comprise an 87 kD polyp ptide. Each domain comprises about 95 amino

acids, with no disulfide bonds. Based on the sequences of the domains, it is apparent that the protein is a member of the cystatin superfamily, comprising repeating cystatin units. Specifically, the domain sequences show high conservation of a QXVXG sequence and the amino terminal region of each domain contains a conserved Gly.

The contiguous domains of the 87 kD potato papain inhibitor polypeptide can be cleaved into individual units comprised of 1, 2, or 3 cystatin domains by the action of trypsin. Other proteases such as chymotrypsin, subtilisin Carlsberg, thermolysin and protease K will also cleave PPI into individual active cystatin units. The domains have a Mr of approximately 10 kD. Each of the domains tested possesses cysteine protease inhibitor activity. From direct amino acid sequencing of several domains and from the predicted amino acid sequences from the DNA sequence, there is close, but not identical, similarity among domains.

Those skilled in the art recognize that such other multidomain cystatins are included within the scope of present invention.

Proteins comprising multidomain cystatin units may be isolated from sources other than potato tubers. For example, naturally occurring multi-domain cystatins may be isolated from various plant sources such as tomato leaflets. Polypeptides of the same size as PPI from potato tubers (87 kD) which react strongly with polyclonal antiserum against tuber PPI are also found in potato leaves. Increased levels of the leaf PPI are inducible by injury to the leaves and levels of up to 2% of the total leaf soluble protein can be achieved. Similar effects are also noted in tomato leaves as the

polypeptide is also recognized by polyclonal antiserum raised against PPI from tubers.

It should be apparent that the present invention, in a specific embodiment, contemplates the use of a midgut-effective plant cystatin comprising one or more of the various cystatin domains. Multi-domain cystatins may be synthesized using repeating combinations of a single domain cystatins or by using different single domain cystatins. For example, the present invention also encompasses inhibitory fragments of the potato papain inhibitor. These fragments include any and all combinations of the eight individual domains of the potato papain inhibitor.

Multi-domain cystatins having more than one type of cystatin subunit are likely to have the ability to control insects resistant to inactivation by any one cystatin subunit. The inhibitors may be rearranged to provide novel protein forms, comprising multi-domain cystatins. It should also be apparent that maintaining a multiplicity of domains will provide, at a molar equivalence level, a cystatin with more efficacious cysteine protease inhibiting properties. Also, the likelihood of pest resistance rapidly developing may be reduced by the inhibitor having a multiplicity of domains with different active site sequences.

Any embodiment which reduces or inhibits the degradation of cystatin in the midgut of the target insect such that the midgut-effective plant cystatin is capable of (or has an increased ability to) be effective in inhibiting the growth of the insect pest is within the scope of this invention. Exemplary embodiments of the present invention include using multi-domain

cystatins and/or co-administering a second inhibitor which inhibits the inactivation of the ingested cystatin.

Multi-domain cystatins may be, or may be prepared to be, of a sufficiently large size that the protein retains its activity for a period sufficient to substantially control the target pest.

Further included herewith are inhibitory fragments of naturally occurring multi-domain cystatins which are also of a sufficiently large size to retains its activity for a period sufficient to substantially control the target pest.

As seen in the Examples section, the multidomain potato papain inhibitor is more resistant to inactivation in the gut than single domains of the 87 kD polypeptide, when ingested by *Diabrotica* larvae. Therefor, given the teachings of the present invention, a skilled artisan may now determine the mode of gut inactivation of the cysteine proteinase inhibitors and include as a synergist an effective amount of a compound which inactivates the cystatin inactivating enzyme(s). For example, when fed to *Diabrotica* larvae, carboxypeptidase inhibitor from potato (Ryan (1974), J. Biol Chem, 249:5495) did not have any deleterious effects on the digestive physiology or growth of the insects. However, the co-administration of potato carboxypeptidase inhibitor and a single cystatin domain derived from the multidomain potato papain inhibitor provided a synergistic effect resulting in growth suppression of *Diabrotica* larvae. Carboxypeptidases from other plant species (e.g., Leary *et al.* (1979), Biochemistry, 18(11):2252-2256) would also be expected

to work as synergists with PPI to control *Diabrotica* larvae.

Using a biochemical assay that monitors the inhibitory activity of a midgut-effective plant cystatin, a skilled artisan may routinely survey plants for proteins with the ability to inhibit cysteine proteases in an *in vitro* assay.—A second screen may be employed in which the cysteine proease inhibitor is applied to the diet of the target insect diet. After ingestion by the insect over 1-2 hours, the midgut is removed and the amount of cysteine proteolytic activity in the gut is assayed using a spectrophotometric, or preferably a sensitive fluorometric or radiometric assay. Those midgut-effective plant cystatins that cause a significant decrease in the amount of digestive cysteine protease activity in the midgut are likely to a have a deleterious effect on the growth of the insect. This can be confirmed by a third screen in which the inhibitor is incorporated in or applied onto the insect diet and the growth and development of the insect monitored by measuring weight gain and mortality.

Once appropriate activity is determined, the amino acid sequence of the midgut-effective plant cystatin, or at least a portion thereof, may be determined by N-terminal sequencing and sequencing of oligopeptides derived by proteolysis. In addition, antisera can be prepared that specifically recognizes the midgut-effective plant cystatin.

It should be understood that, given the present teachings. one may synthesize or isolate substantially pure functional derivatives of naturally-occurring midgut-effective plant cystatins. A "functional

derivative" of the midgut-effective plant cystatin is a compound which possesses a biological activity that is substantially similar to a biological activity of the midgut-effective plant cystatin. The term functional derivative is intended to include "fragments", "effectively homologous variants", or "analogues".

A "fragment" of a molecule is meant to refer to any inhibitory polypeptide subset of a midgut-effective plant cystatin molecule.

An "effectively homologous variant" of a molecule such as the midgut-effective plant cystatin is meant to refer to a molecule substantially similar in sequence and function to either the entire molecule or to a fragment thereof. For purposes of this invention, the structure of one amino acid sequence is effectively homologous to a second amino acid sequence if at least 70 percent, preferably at least 80 percent, and most preferably at least 90 percent of the active portions of the amino acid sequence are identical or equivalent. Generally, the effectively homologous sequences should retain high conservation at the naturally-occurring positions of the amino-terminal Gly residue and the conserved sequence Gln-Xxx-Val-Xxx-Gly. General categories of potentially-equivalent amino acids are set forth below, wherein, amino acids within a group may be substituted for other amino acids in that group: (1) glutamic acid and aspartic acid; (2) lysine, arginine and histidine; (3) alanine. valine, leucine and isoleucine; (4) asparagine and glutamine; (5) threonine and serine; (6) phenylalanine, tyrosine and tryptophan; and (7) glycine and alanine. More importantly and critical to the definition, the function of a second amino acid sequence is effectively homologous to another

amino acid sequence if the second amino acid conforms to a tertiary structure having the capacity to decrease or eliminate the catalytic activity of a digestive cysteine protease.

An "analog" of a molecule such as the midgut-effective plant cystatin is meant to refer to a molecule substantially similar in function to either the entire molecule or a fragment thereof. Thus, provided that two molecules possess a similar activity, they are considered analogs as that term is used herein even if the structure of one of the molecules is not found in the other, or if the sequence of amino acid residues is not identical.

As used herein, the term "substantially pure" is meant to describe the midgut-effective plant cystatin which is homogeneous by one or more purity or homogeneity characteristics. For example, a substantially pure midgut-effective plant cystatin will show constant and reproducible characteristics within standard experimental deviations for parameters such as molecular weight, chromatographic behavior and the like. The term, however, is not meant to exclude artificial or synthetic mixtures of the midgut-effective plant cystatin with other compounds. The term is also not meant to exclude the presence of minor impurities which do not interfere with the biological activity of the midgut-effective plant cystatin and which may be present, for example, due to incomplete purification.

A substantially pure midgut-effective plant cystatin may be isolated from the source in which it naturally exists by any appropriate protein purification technique. Exemplary techniques include chromatographic

techniques, such as gel filtration liquid chromatography, ion exchange chromatography, high performance liquid chromatography, reverse phase chromatography or by use of immunological reagents employing anti-cystatin antibodies.

It is possible to synthesize *in vitro* a midgut-effective plant cystatin from the constituent amino acids (see Merrifield (1963), J. Amer. Chem. Soc., 85:2149-2154; and Solid Phase Peptide Synthesis (1969), (eds.) Stewart and Young). The peptides thus prepared may be isolated and purified by procedures well known in the art (see Current Protocols in Molecular Biology (1989), (eds.) Ausebel, *et al.*) and Sambrook *et al.* (1989), Molecular Cloning: A Laboratory Manual).

Although it is possible to determine and synthesize the entire amino acid sequence of the midgut-effective plant cystatin, it is preferable to isolate the entire sequence of the midgut-effective plant cystatin gene. DNA encoding a midgut-effective plant cystatin may be prepared from chromosomal DNA, cDNA or DNA of synthetic origin by using well-known techniques.

Genomic DNA encoding a midgut-effective plant cystatin may be isolated by standard techniques (Sambrook *et al.* (1989), *supra*). Specifically comprehended as part of this invention are genomic DNA sequences encoding allelic variant forms of the midgut-effective plant cystatin gene, as well as its 5' and 3' flanking regions. It is also possible to use primers and exponentially amplify DNA *in vitro* using sequence specified oligonucleotides by the polymerase chain reaction (PCR) (see Mullis *et al.* (1987), Meth. Enz., 155:335-350; Horton *et al.* (1989), Gene, 77:61; and PCR

Technology: Principles and Applications for DNA Amplification, (ed.) Erlich (1989)).

A DNA isolate encoding a plant non-specific lipid acyl hydrolase may also be obtained from a complementary DNA (cDNA) library. cDNA preparations are ligated into recombinant vectors to form a gene library. Alternatively, the cDNAs may be expressed in a vector such as λgt11 and the library screened using antibodies against the midgut-effective plant cystatin.

A suitable oligonucleotide or set of oligonucleotides may be used, by techniques well known in the art, to screen the genomic DNA or cDNA libraries. To facilitate the detection of the desired sequence, the oligonucleotide probe may be labeled with any material having a detectable physical or chemical property. General procedures for isolating, purifying and sequencing the desired sequences are well known in the art (see Current Protocols in Molecular Biology (1989), *supra*; and Sambrook *et al.* (1989), *supra)*.

An alternative way of obtaining a genetic sequence which is capable of encoding the midgut-effective plant cystatin is to prepare it by oligonucleotide synthesis, after the gene sequence of interest is determined (see Caruthers (1983), In:Methodology of DNA and RNA, (ed.) Weissman); Beaucage *et al.* (1981), Tetrahedron Letters, 22:1859-1962). A series of oligonucleotides may be synthesized in order to provide a series of overlapping fragments which when annealed and ligated will produce both strands of the gene. These fragments are then annealed and ligated together using well known techniques (see Sambrook *et al.* (1982), *supra*). Alternatively, the gene may be produced

by synthesizing a primer having a so-called "wagging tail", that does not hybridize with the target DNA; thereafter, the genomic sequences are amplified and spliced together by overlap extension (see Horton *et al.* (1989), Gene, 77:61-68). The resulting DNA fragment with the predicted size is isolated by electrophoresis and ligated into a suitable cloning vector for amplification and further manipulation (see Mullis *et al.* (1987), *supra*; and PCR Technology: Principles and Applications for DNA Amplification, *supra*).

Of course, one may incorporate modifications into the isolated sequences including the addition, deletion, or nonconservative substitution of a limited number of various nucleotides or the conservative substitution of many nucleotides, provided that the proper reading frame is maintained. Translational stop and start signals are added at the appropriate points, and sequences to create convenient cloning sites are added to the ends. Exemplary techniques for modifying oligonucleotide sequences include using polynucleotide-mediated, site-directed mutagenesis (see Zoller *et al.* (1984), DNA, 3:479-488); Higuchi *et al.* (1988), Nucl. Acids Res., 16:7351-7367; Ho *et al.* (1989), Gene, *supra*; Horton *et al.* (1989), *supra*; and PCR Technology: Principles and Applications for DNA Amplification, (ed.) Erlich (1989)).

In order to further characterize such genetic sequences, it is desirable to introduce the sequence into a suitable host to express the proteins which these sequences encode, and confirm that they possess characteristics of midgut-effective plant cystatin.

T chniques for such manipulations are well known in the art and disclosed by Sambrook *et al.* (1989), *supra*.

Vectors are available or can be readily prepared for transformation of viruses, prokaryotic or eukaryotic cells. In general, plasmid or viral vectors should contain all the DNA control sequences necessary for both maintenance and expression of a heterologous DNA sequence in a given host. Such control sequences generally include a promoter sequence, a transcriptional start or leader sequence, a DNA sequence coding for translation start-signal codon, a translation terminator codon, and a DNA sequence coding for a 3' non-translated region containing containing a signals controlling termination of RNA synthesis and/or messenger RNA modification. Finally, the vectors should desirably have a marker gene that is capable of providing a phenotypical property which allows for identification of host cells containing the vector, and an intron in the 5' untranslated region, e.g., intron 1 from the maize alcohol dehydrogenase gene that enhances the steady state levels of mRNA.

Exemplary host cells include prokaryotic and eukaryotic strains. The appropriate procedure to transform a selected host cell may be chosen in accordance with the host cell used. Based on the experience to date, there appears to be little difference in the expression of genes, once inserted into cells, attributable to the method of transformation itself.

Conventional technologies for introducing biological material into host cells include electroporation [see Shigekawa and Dower (1988),

Biotechniques, 6:742; Miller, *et al.* (1988), Proc. Natl. Acad. Sci. USA, 85:856-860; and Powell, *et al* (1988), Appl. Environ. Microbiol., 54:655-660]; direct DNA uptake mechanisms [see Mandel and Higa (1972), J. Mol. Biol., 53:159-162; Dityatkin, *et al.* (1972), Biochimica et Biophysica Acta, 281:319-323; Wigler, *et al.* (1979), Cell, 16:77; and Uchimiya, *et al.* (1982), In: Proc. 5th Intl. Cong. Plant Tissue and Cell Culture, A. Fujiwara (ed.), Jap. Assoc. for Plant Tissue Culture, Tokyo, pp. 507--508]; fusion mechanisms [see Uchidaz, *et al.* (1980), In: Introduction of Macromolecules Into Viable Mammalian Cells, C. Baserga, G. Crose, and G. Rovera (eds.) Wistar Symposium Series, Vol. 1, A. R. Liss Inc., NY, pp. 169-185]; infectious agents [see Fraley, *et al.* (1986), CRC Crit. Rev. Plant Sci., 4:1-46); and Anderson (1984), Science, 226:401-409]; microinjection mechanisms [see Crossway, *et al.* (1986), Mol. Gen. Genet., 202:179-185]; and high velocity projectile mechanisms [see EPO 0 405 696].

Transformants are isolated in accordance with conventional methods, usually employing a selection technique, which allows for selection of the desired organism as against unmodified organisms. Generally, after being transformed, the host cells are grown for about 48 hours to allow for expression of marker genes. The cells are then placed in selective and/or screenable media, where untransformed cells are distinguished from transformed cells, either by death or a biochemical property. The selected cells can be screened for expression of the midgut-effective plant cystatin by assay techniques such as immunoblot analysis, enzyme-linked immunosorbent assay, radioimmunoassay, or fluorescence-activated cell sorter analysis,

immunohistochemistry and the like. The transformed tissues are then tested for insect controlling activity.

A host cell may be transformed to provide a source from which significant quantities of the vector containing the gene of interest can be isolated for subsequent introduction into the desired host cells or for which significant quantities of the protein may be expressed and isolated. Exemplary recombinant host cells include unicellullar prokaryotic and eukaryotic strains. Prokaryotic microbes that may be used as hosts include *Escherichia coli*, and other *Enterobacteriaceae*, *Bacilli*, and various *Pseudomonas*. Common eukaryotic microbes include *Sacchromyces cerevisiae* and *Pichia pastoris*. Common higher eukaryotic host cells include Sp2/0 or CHO cells. Another preferred host is insect cells, for example *Drosophila* larvae, in which the vector contains the *Drosophila* alcohol dehydrogenase promoter. Alternatively, baculovirus vectors, e.g., *Autographa californica* nuclear polyhedrosis virus (see Miller *et al.* (1983), Science, 219:715-721) may be engineered to express large amounts of the midgut-effective plant cystatin in cultured insects cells (see Andrews *et al.* (1988), Biochem J., 252:199-206.

The present invention provides an agricultural composition for application to plants or parts thereof which are susceptible to infestation by insects having digestive cysteine proteases, said agricultural composition comprising a midgut-effective plant cystatin. Often the agricultural composition will contain an agriculturally acceptable carrier. By the term "agriculturally acceptable carrier" is meant a substance which may be used to dissolve, disperse or diffuse an active compound in the composition without

impairing the effectiveness of the compound and which by itself has no detrimental effect on the soil, equipment, crops or agronomic environment.

The agricultural compositions may be applied in a wide variety of forms including powders, crystals, suspensions, dusts, pellets, granules, encapsulations, microencapsulations, aerosols, solutions, gels or other dispersions. In addition to appropriate liquid or solid carriers, compositions may include adjuvants, such as emulsifying and wetting agents, spreading agents, dispersing agents, adhesives or agents which stimulate insect feeding according to conventional agricultural practices. Adjuvants for the formulation of insecticides are well known to those skilled in the art.

The concentration of midgut-effective plant cystatin will vary widely depending upon the nature of the particular formulation, particularly whether it is a concentrate or it is to be used directly. The midgut-effective plant cystatin generally will be present in at least 1 percent by weight and may be up to 100 percent by weight.

The presentation of the agricultural composition may be achieved by external application either directly or in the vicinity of the plants or plant parts. The agricultural compositions may be applied to the environment of the insect pest(s), e.g., plants, soil or water, by spraying, dusting, sprinkling, or the like.

The present invention further contemplates using recombinant hosts (e.g., microbial hosts and insect viruses) transformed with a gene encoding the

midgut-effective plant cystatin and applied on or near a selected plant or plant part susceptible to attack by a target insect. The hosts are selected capable of colonizing a plant tissue susceptible to insect infestation or of being applied as dead or non-viable cells containing the midgut-effective plant cystatin. Microbial hosts of particular interest will be the prokaryotes and the lower eukaryotes, such as fungi.

Characteristics of microbial hosts for encapsulating a midgut-effective plant cystatin include protective qualities for the protein, such as thick cell walls, pigmentation, and intracellular packaging or formation of inclusion bodies; leaf affinity; lack of mammalian toxicity; attractiveness to pests for ingestion; ease of killing and fixing without damage to the midgut-effective plant cystatin; and the ability to be treated to prolong the activity of the midgut-effective plant cystatin. Characteristics of microbial hosts for colonizing a plant include non-phytotoxicity; ease of introducing a genetic sequence encoding a midgut-effective plant cystatin, availability of expression systems, efficiency of expression and stability of the insecticide in the host.

Illustrative prokaryotes, both Gram-negative and -positive, include *Enterobacteriaceae*, such as *Escherichia*; *Bacillaceae*; *Rhizoboceae*, such as *Rhizobium and Rhizobacter*; *Spirillaceae* (such as photobacterium), *Zymomonas*, *Serratia*, *Aeromonas*, *Vibrio*, *Desulfovibrio*, *Spirillum*; *Lactobacillaceae*; *Pseudomonadaceae* (such as *Pseudomonas* and *Acetobacter*); *Azotobacteraceae* and *Nitrobacteraceae*. Among eukaryotes are fungi (such as *Phycomycetes* and *Ascomycetes*), which includes yeast (such as *Saccharomyces* and *Schizosaccharomyces*); and *Basidiomycetes*

yeast (such as *Rhodotorula*, *Aureobasidium*, *Sporobolomyces*) and the like.

The present invention also contemplates the use of a baculovirus containing a gene encoding a midgut-effective plant cystatin. Baculoviruses including those that infect *Heliothis virescens* (cotton bollworm), *Orgyla pseudotsugata* (Douglas fir tussock moth), *Lymantria dispar* (gypsy moth), *Autographica californica* (alfalfa looper), *Neodiprion sertifer* (European pine fly) and *Laspeyresia pomonella* (coddling moth) have been registered and used as pesticides (see US 4,745,051 and EP 175 852).

The recombinant host may be formulated in a variety of ways. It may be employed in wettable powders, granules or dusts, or by mixing with various inert materials, such as inorganic minerals (phyllosilicates, carbonates, sulfates, phosphates, and the like) or botanical materials (powdered corncobs, rice hulls, walnut shells, and the like). The formulations may include spreader-sticker adjuvants, stabilizing agents, other insecticidal additives surfactants, and bacterial nutrients or other agents to enhance growth or stabilize bacterial cells. Liquid formulations may be aqueous-based or non-aqueous and employed as foams, gels, suspensions, emulsifiable concentrates, or the like. The ingredients may include rheological agents, surfactants, emulsifiers, dispersants, or polymers.

Alternatively, the midgut-effective plant cystatin can be incorporated into the tissues of a susceptible plant so that in the course of infesting the plant the insect consume insect-controlling amounts of the selected midgut-effective plant cystatin. One

method of doing this is to incorporate the midgut-effective plant cystatin in a non-phytotoxic vehicle which is adapted for systemic administration to the susceptible plants. However, since the genes which code for midgut-effective plant cystatin may be isolated, the invention contemplates, in a preferred embodiment, transgenic plants which are capable of biologically synthesizing midgut-effective plant cystatin to provide the plants with a new, or an additional, mechanism of protection against attack by insects.

The invention provides methods of imparting resistance to insect infestation by insects having digestive cysteine proteases to plants of a susceptible taxon, comprising: (a) culturing cells or tissues from at least one plant from the taxon; (b) introducing into the cells of the cell or tissue culture a structural gene encoding a midgut-effective plant cystatins operably linked to plant regulatory sequences which cause expression of the midgut-effective plant cystatin gene in the cells, and (c) regenerating insect-resistant whole plants from the cell or tissues culture.

Obviously, the expression of uniquely high quantities of midgut-effective plant cystatins may be deleterious to the plant itself. The use of a signal sequence to secrete or sequester in a selected organelle allows the protein to be in a metabolically inert location until released in the gut environment of an insect pathogen. Moreover, some proteins are accumulated to higher levels in transgenic plants when they are secreted from the cells, rather than stored in the cytosol (Hiatt, *et al.* (1989), Nature, 342:76-78).

The DNA sequence will gen rally be one which originates from, or has substantial sequence homology to a midgut-effective plant cystatin, originating from a plant of a species different from that of the target organism. However, when the DNA sequence is one which originates from, or has substantial sequence homology to a midgut-effective plant cystatin originating from, a plant of the same species as that of the target plant, such sequence may be expressed in significantly greater amounts.

In order to optimize the transcriptional and translational efficiency of such systems, it is possible to examine the frequency of codon usage and determine which codons are, in essence, preferred within the transcriptional and translational systems normally present in that plant. Using such preferred usage codons, it is possible to construct a protein coding sequence which may result in a significantly enhanced level of transcriptional and translational efficiency of the midgut-effective plant cystatin gene compared to what would be achieved by taking the coding sequence directly in an ummodifed form of the donor plant.

Generally, the insertion of heterologous genes appears to be random using any transformation technique; however, technology currently exists for producing plants with site specific recombination of DNA into plant cells (see WO/9109957). The activity of the foreign gene inserted into plant cells is dependent upon the expression characteristics of the individual inserted genes, resulting from control regions (promoters, polyadenylation regions, enhancers, etc.)

and from the influence of endog nous plant DNA adjacent the chimeric insert and by the copy number.

The promoter selected should be capable of causing sufficient expression to result in the production of an insect controlling amount of protein. Suitable promoters may include both those which are derived from a gene which is naturally expressed in plants and synthetic promoter sequences which may include redundant or heterologous enhancer sequences. In cases where the sequence is derived from a plant source, one can use the 5' and 3' non-translated region naturally associated with the particular gene. A number of promoters which are active in plant cells include the nopaline synthase, octopine synthase and mannopine synthase promoters from the tumor-inducing plasmids of *Agrobacterium tumefaciens*.

In species which produce a midgut-effective plant cystatin but in lower than insecticidal amounts, it may be preferable to overexpress the midgut-effective plant cystatins in the same plant, and even tissue, from which it was derived, wherein the midgut-effective plant cystatins is expressed at significantly greater levels than normally found. By significantly greater levels is meant the production of the midgut-effective plant cystatins at levels at least 50% greater than normally found in untransformed plants of the same species. Accordingly, the present invention contemplates constitutive promoters such that the transformed plant has increased tolerance to insect pests. Examples of constitutive promoters include the CaMV 19S and 35S promoters (JP 63287485), ubiquitin promoter, the rice actin promoter (WO 9109948).

In species which produce a native midgut-effective plant cystatin which is not produced in or not distributed to tissues which are normally infested with the insects, a tissue specific promoter can be used to provide localized expression of or overproduction of the midgut-effective plant cystatins. Examples of tissue specific promoters include the root specific promoters such as maize metallothionein (EP 452269), the root specific promoter (WO/9113992) the plant seed storage body promoter (WO/9113993), and the alcohol dehydrogenase-1 promoter. Promoters known to be light inducible include the promoter of the gene encoding the small subunit (ss) of the ribulose-1,5,-bisphosphate carboxylase from soybean and the promoter of the gene encoding the chlorophyll a/b binding protein in greening leaves (Coruzzi *et al.*, (1983), J. Biol. Chem., 258:1399; and Dunsmuir, *et al.* (1983), J. Molecular and App. Gen., 2:285).

Finally, a wound or pathogen inducible promoter can be used to provide expression of the midgut-effective plant cystatins when a tissue is attacked by a plant pest. Examples of wound or pathogen inducible promoters include the proteinase inhibitor II promoter.

Suitable vectors for transforming plant tissue and protoplasts have been described in the literature and are set forth herein (see deFrammond *et al.* (1983), Biotechnology, 1:262; An *et al.* (1985), EMBO J. 4:277; Potrykus *et al.* (1985), Mol. Gen. Genet. 199:183; Rothstein *et al.* (1987), Gene, 53:153; WO 90/08829 and WO 84/02913; and, in a preferred embodiment, pDAB219Δ-Not (as described in the Examples). It is not necessary in practice that the vector harboring the selectable marker gene also contain the gene of interest. Rather, co-

transformation of such vectors may be used to transform plant cells.

The appropriate procedure to produce mature transgenic plants may be chosen in accordance with the plant species used. Regeneration varies from species to species of plants. Efficient regeneration will depend upon the medium, on the genotype and on the history of the culture. Once whole plants have been obtained, they can be sexually or clonally reproduced in such a manner that at least one copy of the sequence is present in the cells of the progeny of the reproduction. Such procedures may be chosen in accordance with the plant species used.

Mature plants, grown from the transformed plant cells, may be selfed to produce an inbred plant. In diploid plants, typically one parent may be transformed and the other parent may be the wild type. The parent will be crossed to form first generation hybrids (F_1), which are selfed to produced second generation hybrids (F_2). F_2 hybrids with the genetic makeup of midgut-effective plant cystatin/midgut-effective plant cystatin are chosen and selfed to produce an inbred plant.

Conventional plant breeding methods can be used to transfer the midgut-effective plant cystatin structural gene via crossing and backcrossing. Such methods comprise the further steps of (a) sexually crossing the insect-resistant plant with a plant from the insect-susceptible variety; (b) recovering reproductive material from the progeny of the cross; and (c) growing insect-resistant plants from the reproductive material. Where desirable or necessary, the agronomic characteristics of the susceptible variety

can be substantially preserved by expanding this method to include the further steps of repetitively (d) backcrossing the insect-resistant progeny with insect-susceptible plants from the susceptible variety; and (e) selecting for expression of insect resitance (or an associated marker gene) among the progeny of the backcross, until the desired percentage of the characteristics of the susceptible variety are present in the progeny along with the gene imparting insect resistance. Subsequently, the inbreds according to this invention may be crossed with another inbred line to produce the hybrid.

The present invention further contemplates using, with the midgut-effective plant cystatin, adjuvants, chemical or biological additives in an effort to expand the spectrum of target pests, to extend the duration of effectiveness of the midgut-effective plant cystatin or to help stabilize the agricultural composition of the midgut-effective plant cystatin.

Exemplary potentiators would include lectins, amphipathic proteins or complementary proteinase inhibitors. For example, the presence of more than one defensive protein, in the presence of other defensive proteins, may have an important role in the plant defense against insect attacks. It is known that *Hemiptera* and *Coleoptera* insects developed alternative pathways of protein digestion of foods containing high levels of certain proteinase inhibitors. It may be advantagous to include inhibitors from families such as Kunitz-type inhibitors, Bowman-Birk inhibitors, Barley Trypsin inhibitors, Potato inhibitors I and II, Squash inhibitors, Ragi 1-2/Maize bifunctional inhibitors, carb xyp ptidase A and B inhibitors and aspartyl

proteinase inhibitors (s e Ryan (1990), Annu. Rev. Phtyopathol., 28:425-49.

The present invention contemplates protecting any plant of a taxon which is susceptible to infestation and damage by insects having digestive cysteine proteases, especially one or more of corn rootworm, Mexican been beetle, red flour beetle, confused flour beetle, cowpea beetle, boll weevil, Colorado potato beetle, three-lined potato beetle, rice weevil, maize weevil, granary weevil, lesser grain borer, flea beetles, Egyptian alfalfa weevil, bean weevil, yellow mealworm, asparagus beetle and squash bug. By the term "taxon" herein is meant a unit a botanical classification of genus or lower. It thus includes genus, species, cultivars, varieties, variants and other minor taxonomic groups which lack a consistent nomenclature.

Exemplary plants include maize, sorghum, tomato, potato, cotton, soybean, dry beans, rape, alfalfa, asparagus and sweet potato. However, it is not to be construed as limiting, inasmuch as these insects may infest certain other crops. Thus, the methods of the invention are readily applicable to numerous plant species, if they are found to be susceptible to the plant species listed hereinabove, including without limitation, species from the genera *Medicago*, *Trifolium*, *Vigna*, *Citrus*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Capsicum*, *Lycopersicon*, *Nicotinia*, *Solanum*, *Helianthus*, *Bromus*, *Asparagus*, *Panicum*, *Pennisetum*, *Cucumis*, *Glycine*, *Lolium*, *Triticum* and *Zea*.

Examples

The present invention is illustrated in further detail by the following examples. The examples are for the purposes of illustration only, and are not to be construed as limiting the scope of the present invention. All parts and percentages are by weight unless otherwise specifically noted. All DNA sequences are given in the conventional 5' to 3' direction. All amino acid sequences are given in conventional amino-terminus to carboxylic acid terminus direction.

Example 1: Purification and characterization of potato papain inhibitor

A. Purification

Potato papain inhibitor (PPI) was purified from the skin of market-purchased potato tubers essentially as described in the literature (Rodis and Hoff, (1984), *supra*). Typical yields were 10-50 mg pure PPI from the peel from 10 lb of tubers. After purification, the protein was homogeneous as determined by sodium dodecyl sulfate-polyacracylamide gel electrophoresis and by ion exchange and reverse-phase high performance liquid chromatography.

B. Fragmentation of PPI by limited proteolysis

PPI was fragmented into smaller polypeptides that retained the ability to inhibit cysteine proteases by incubating PPI with trypsin (treated with L-1-chloro-3-(4-tosylamido)-7-amino-2-heptanone chloride) at a PPI:trypsin ratio of 20:1 (weight basis) for 2 hours at 37°C in 50 mM Tris buffer, pH 7.5. This resulted in a stable family of polypeptides of ~10 kD and polypeptides

of ~32 kD. Fragmentation of PPI was also achieved using other proteases such as chymotrypsin, subtilisin Carlsberg, protease K and thermolysin using the same conditions. In these instances, a 22 kD (rather than 32 kD) species was released in addition to the 10 kD species.

Trypsin-treated PPI was fractionated by size exclusion chromatography using a Superose 12 column (Pharmacia LKB, Piscataway, NJ). This resulted in two peaks, the first of which contains the ~32 kD species and the second, the ~10 kD species. These were further purified by ion exchange chromatography using a Mono Q HR™ 5/5 column equilibrated in 20 mM Tris pH 7.5 and developed with a 0-150 mM NaCl gradient over 30 minutes at a flow rate of 1 ml/minutes. The resulting fragments (two ~32 kD species and six ~10 kD species) are all potent inhibitors of papain and of *Diabrotica* gut proteases (see Example 2).

Trypsin-digested PPI was also separated by reverse phase chromatography using a Vydac C4 column equilibrated in 0.1% trifluoroacetic acid and developed with a 0-60% acetonitrile gradient. This resulted in five peaks containing the ~10 kD polypeptides (named PPI-10K-1 through 5) and two peaks containing the ~32 kD polypeptides (named PPI-32K and PPI-33K).

C. Stoichiometry and inhibition constants for PPI and the tryptic fragments

The inhibitory activity of the polypeptide was assayed by monitoring the inhibition of papain in the presence of varying amounts of PPI. Papain activity was measured by methods as described and referenced in

"Proteolytic Enzymes: a practical approach" (ed.) R. J. Benyon and J. S. Bond (1989), IRL Press, NY. Substrates were azocasein, benzoyl-arginine para-nitroanilide (BAPNA) (spectrophotometric assay) or Z-phe-arg-NMec (where "Z" = benzoyloxycarbonyl, and NMec = 7-amido-4-methylcoumarin) (fluorometric assay).

Titrations of papain with PPI using BAPNA as substrate and spectrophotometrically monitoring the release of para-nitrophenol at 440 nm showed that 28 pmoles PPI stoichiometrically inhibited 217 pmoles papain. Typically 5 µg papain was titrated with PPI in 0.1 ml 0.1 M sodium acetate, pH 6; 4 mM dithiothreitol 2 mM sodium EDTA. Reactions were initiated with 0.1 ml BAPNA and the absorbance at 410 recorded after 1 hour using a microplate reader. The stoichiometries obtained indicate that there are 8 binding sites for papain per PPI molecule. This stoichiometry is unchanged after trypsin treatment to fragment the protein into the 10 kD and 32 kD polypeptides demonstrating that the ability to inhibit cysteine protease is maintained even after limited proteolysis.

The inhibition constants (Ki values) for PPI-87K, PPI-10K and PPI-32K against papain were determined by the sensitive fluorometric assay using Z-Phe-Arg-NMec (adapted from Barrett and Kirschke (1981), Methods Enzymol., **80**, 540-541). Initially, titrations of papain (50 ng) were performed with various concentrations of PPI (10-100 ng) at 20 µM final substrate concentration to obtain inhibition profiles of each polypeptide. Reactions were started by addition of Z-Phe-Arg-NMec (0.2 ml) using a Costar multipipettor to give a final volume of 0.3 ml. Hydrophobic interactions of PPI with the microtiter plate required the addition of 0.01%

Triton X-100 (v/v) to obtain data comparable to that seen at higher papain/inhibitor concentrations using the less sensitive BAPNA assay. The presence or absence of Triton had no apparent effect on calculated Ki values. Data were collected using a Fluroskan II™ fluorescent microtiter plate reader attached to a Biometallics™ data collection software system having kinetic capabilities. Excitation and emission wavelength were 380 and 460 nm, respectively. Kinetic assays from which Ki data were derived were typically performed at inhibitor concentrations necessary to give levels of 30, 50, and 70% inhibition of papain, and substrate concentrations ranged from 1 to 20 μM. The Ki values for PPI-87K, PPI-32K and PPI-10K were 0.1 (±0.01), 0.7 (±0.06) and 0.5 (±0.02) nM, respectively, indicating that the inhibitory potency of the fragments is similar to that of the parent molecule.

D. Amino acid sequences of PPI and tryptic fragments.

The PPI fragments purified by reverse-phase chromatography were subjected to N-terminal amino acid sequencing using an ABI 477 liquid phase sequencer with an online 120A PTH amino acid analyzer. Extended sequence information was obtained for PPI-10K-4 and PPI-10K-5. The parent 87 kD polypeptide (PPI-87K) was also sequenced, but was resistant to Edman degradation. PPI-10K-2 was also resistant to N-terminal sequence analysis and so is presumably the N-terminal fragment derived from PPI-87K. PPI-10K-2 was denatured in urea and further digested by trypsin and the resulting peptides separated by reverse-phase chromatography. Three of the purified oligopeptide fragments were sequenced: PPI-10K-2-T77, PPI-10K-2-T51 and PPI-10K-2-T32.

The amino acid sequences are shown in Fig 3. There is ext nsive homology between the sequences indicating that the fragments are all related. The extended sequence data obtained from PPI-10K-5 was compared to that of hen egg cystatin and rice cystatin (Fig. 4). These data indicate that PPI consists of 8 domains that are all closely related and which are all members of the cystatin family of cysteine protease inhibitors.

Example 2: Inhibition of *Diabrotica* larval growth by PPI

PPI is functional as a protease inhibitor either as an intact 87 kD polypeptide or after proteolytic cleavage into smaller one, two or three domain fragments.

The effect of PPI on the larval growth of *Diabrotica spp.* was monitored as follows: 0.03 ml of purified PPI solutions (in water) was applied to the surface of 0.25 ml artificial diet (adapted from Rose and McCabe, J. Econ. Entomol., 66:398-400, 1973) in 24 well plates and allowed to air dry in a sterile flow hood. The wells were then infested with single, neonate southern corn rootworm (SCR, *Diabrotica undecimpunctata howardi*) hatched from sterilized eggs or with single, preweighed second instar western corn rootworm (WCR, *Diabrotica virgifera virgifera*). The plates were then p[illegible]ed in sterilized, sealed plastic containers and put in a humidified growth chamber maintained at 25°C for 6 days (SCR) or 3.5 days (WCR) prior to final weighing.

A. Effect on SCR neonate larvae

Increasing concentrations of purified PPI were used in feeding studies with neonate SCR as described

above. PPI caused a dose-dependent inhibition of larval growth with 50% inhibition occurring at approximately 0.2 mg/g diet (Figure 5). Maximal inhibition observed was 79% at 2.0 mg/g diet. These data indicate that PPI is an effective inhibitor of SCR neonate larval growth.

B. Effect on second-instar WCR and SCR

WCR larvae cannot be bioassayed as neonates due to their inability to grow on artificial diet. However, second instar larvae will grow and develop on artificial diet and feeding studies were conducted as described above. PPI caused a dose-dependent inhibition of WCR larval growth with 50% inhibition occurring at approximately 0.25 mg PPI/g diet (Figure 6). At the highest dose tested (1.0 mg/g diet), PPI essentially halted growth. Similar experiments with second instar SCR showed a reduced effect of PPI on larval growth (50% inhibition at ~1.0 mg PPI/g diet) in comparison to neonate larvae. These data show that PPI is an effective inhibitor of WCR larval growth. In addition, the stage specific effects of PPI in SCR suggest that neonate WCR larvae may be more sensitive to the effects of PPI than neonate SCR.

Example 3: Synergistic action of PPI-10K in combination with a carboxypeptidase inhibitor

As demonstrated in Example 1, tryptic digestion of the intact PPI molecule produces ~10 and ~32 kD fragments which are effective *in vitro* inhibitors of papain. Intact PPI and the isolated PPI ~10 kD fragments also inhibit proteolysis by *Diabrotica* rootworm gut extracts using the same assay system. However, these fragments have little or no growth inhibitory effect in

feeding assays using either neonate SCR and second instar WCR (Table 1). To resolve this difference, a sensitive assay was developed to directly monitor the cysteine proteolytic activity in the gut juice of larvae after ingestion of intact PPI or PPI fragments.

0.03 ml of either PPI-87K or PPI-10K (obtained from trypsin digestion of 87 kD) in water was applied at a final concentration of 0.5 mg/ml of diet to the surface of 0.25 ml artificial diet as described above. The diet from 10 wells was then removed and place in a sterile petri dish. Twenty two instar SCR (reared on corn roots) were then placed on the diet and allowed to feed for 3 hours. After feeding the alimentary canal from each larva was removed and placed in individual chilled Eppendorf tubes containing 50 µl 50 mM sodium acetate buffer, pH 6.0 + 4 mM EDTA + 8 mM dithiothreitol and vortexed vigorously. Five microliters of the extract from each tube was then added to 0.095 ml of the same buffer in a 96-well plate and allowed to sit for 10 minutes. 0.2 ml of Z-Phe-Arg-NMec were added to the diluted extract using a Costar 96-well pipettor to give 20 µM final concentration. Data were collected via a Fluroskan II™ fluorescent microtiter plate reader attached to a Biometallics™ data collection software system having kinetic capabilities. Excitation and emission wavelength were 380 and 460, respectively. Initial rates of substrate cleavage were expressed as fluorescence units/minute.

These data show that ingestion of PPI-10K resulted in relatively little inhibition of cysteine protease activity in the rootworm larval midgut. In

Table 1: Comparison of cysteine protease activity in rootworm larval midguts after feeding PPI derivatives:

Sample	Activity	Control (fluorescence units/min)
Control	297.1± 36.5	100
PPI-87K	70.8± 14.7	24
PPI-10K	242.5± 32.9	82

contrast, ingestion of PPI-87K effectively inhibits cysteine protease activity within the gut.

The effect of potato carboxypeptidase inhibitor in the presence of PPI-10K was determined in insect feeding assays. These data are shown in Table 2.

Table 2: Effect of PPI tryptic fragments on growth of neonate southern corn rootworm larvae:

TREATMENT (mg/g diet)	LARVAL WEIGHT (mg)
Control	3.50 ± 0.11*
PPI-10 (1.0)	3.40 ± 0.21
PPI-10 (0.25)	3.09 ± 0.19
PPI-32 (0.25)	3.70 ± 0.28
CPI (0.25)	3.60 ± 0.14
PPI-10 (0.25) + CPI (0.25)	2.15 ± 0.15

* - Values are the mean ± SEM from 5 separate experiments

Neither potato carboxy-peptidase inhibitor or PPI-10K alone (each at 0.25 mg/g diet) had any effect on rootworm larval growth. However, when the two proteins were combined in the diet, a significant reduction in growth was seen comparable to that seen with an equal concentration of intact PPI. Pepstatin, a carboxyl protease inhibitor had a similar but smaller effect.

No significant effect of PPI-10K was seen in feeding assays of 2nd instar WCR (Table 3). Additionally, no significant effect of PPI-10K with carboxypeptidase inhibitor was seen in feeding assays of 2nd instar WCR (data not shown).

Table 3: Effect of PPI tryptic fragments on growth of second instar western corn rootworm larvae

TREATMENT (mg/g diet)	WEIGHT GAIN (mg)
Control	7.65 ± 0.35*
PPI - 8 + 32kD (0.25)	5.99 ± 0.41
PPI - 87 kD (0.25)	2.81 ± 0.44

* Values are the mean ± SEM from 2 separate experiments

Example 4 - Eff ct of PPI in combination with a carboxyl protease inhibitor

PPI inhibited 60-90% of rootworm gut proteolysis depending on the type of protein used as a substrate in the assay. The remaining proteolytic activity was effectively inhibited by pepstatin, a carboxyl protease inhibitor. This was shown as follows: diluted SCR gut extract and the appropriate inhibitors (PPI was used at 40 µg/ml final concentration pepstatin at 8 µg/ml final concentration) were combined in a total volume of 0.05 ml 100 mM sodium acetate buffer pH 6 + 5 mM EDTA + 5 mM dithiothreitol. The assays were initiated with 20 µl 0.5% fluorescein isothiocyanate (FITC)-linked hemoglobin or casein prepared as described in Twining (Anal. Biochem, 143:30) and quenched after 3 hours at 37°C with 75 µl 12.5% trichloracetic acid. After 30 minutes at 4°C, samples were centrifuged and the fluorescence of the supernatant determined after mixing with 110 µl M Tris pH. Results are given as percent (%) inhibition relative to a control containing no inhibitors.

Table 4: Effect of PPI and pepstatin on *in vitro* proteolysis using SCR larval gut extracts

Inhibitor	Substrate	Inhibition
PPI	FITC casein	91.4
PPI	FITC hemoglobin	59.0
Pepstatin	FITC casein	39.4
Pepstatin	FITC hemoglobin	48.6
PPI + pepstatin	FITC casein	97.6
PPI + pepstatin	FITC hemoglobin	99.2

Similar results were obtained with WCR gut extracts, using azoalbumin and azocasein as substrates. The extent of inhibition of larval gut proteolysis by PPI varies according to the protein substrate in the order casein>albumin>hemoglobin. Inhibition by pepstatin has the reverse order with most inhibition observed with hemoglobin as substrate. Although PPI alone consistently inhibits a larger proportion of rootworm gut proteolysis than pepstatin alone, neither inhibitor gives complete inhibition. However, the combination of both inhibitors gives almost complete inhibition of proteolysis for all protein substrates. The combination of an aspartate protease inhibitor with PPI in the diet gives enhanced inhibition of *Diabrotica* larval growth, when assayed as described in Example 2.

Example 5: Construction of plant expression plasmids

The plasmid pDAB219Δ-Not represents a dual purpose vector containing two genes, each under the control of a promotor expressed in callus tissue. The

first gene, a screenable marker, is a modified beta-glucuronidase (*gus*) gene from *Escherichia coli* under the translational control of the CaMV 35S promoter. Plant transcription and poly-adenylation addition signals are supplied by sequences derived from a nopaline synthase gene. The second gene, *bar*, is a selectable marker which codes for phosphinothricine acetyl transferase and is derived from *Streptomyces hygroscopicus*. This gene is also under the regulation of the CaMV 35S promoter and nopaline synthase transcription termination poly-adenylation sequences. The *gus* gene allows for the rapid analysis of expression using commercially available fluorometric or histochemical assays. The expression of the *bar* gene confers resistance to the herbicide Basta™ (Hoechst), which contains as an active ingredient bialaphos, thus, imparting a selective advantage to transformed cells under selection pressure. The sequences derived from Cauliflower Mosaic Virus (CaMV) represent the Cabb S strain. They are available as the MCASTRAS sequence of GenBank, and published by Franck *et al.* (1980), Cell, 21:285-294.

A. Plasmids utilizing the 35S promoter and the Agrobacterium Nos Poly A sequences

The starting material for is plasmid pBI221, purchased from CLONTECH (Palo Alto, CA). This plasmid contains a modified copy of the CaMV 35S promoter, as described in Bevan *et al.* (1985), EMBO J., 4:1921-1926; Baulcombe *et al.* (1986), Nature, 321:446-449; Jefferson *et al.* (1987), EMBO J., 6:3901-3907; and Jefferson (1987), Plant Molec. Biol. Reporter, 5:387-40. Beginning at the 3' end of the *Pst*I site of pUC 19 (Yanisch-Perron *et al.* (1985), Gene, 33:103-119), and reading on the same strand as that which encodes the Lac Z gene of pUC 19,

the sequence is compris d of the linker nucleotides GTCCCC, followed by CaMV nucleotides 6605 to 7439, followed by the linker sequence GGGGACTCTAGAGGATCCCCGGGTGGTCAGTCCCTT, wherein the underlined bases represent the *Bam*HI recognition sequence. These bases are then followed by 1809 base pairs (bp) comprising the coding sequence of the *Escherichia coli* uidA gene, which encodes the b-glucuronidase (GUS) protein, and 44 bp of 3' flanking bases that are derived from the *E. coli* genome (Jefferson, *et al.* (1986), Proc. Natl. Acad. Sci., 83:8447-8451), followed by the *Sst*I linker sequence GAGCTC, which is then followed by the linker sequence GAATTTCCCC. These bases are followed by the RNA transcription termination/polyadenylation signal sequences derived from the *Agrobacterium tumefaciens* nopaline synthase (*Nos*) gene, and comprise the 256 bp *Sau*3AI fragment corresponding to nucleotides 1298 to 1554 of DePicker *et al.* (1982), (J. Molec. Appl. Genet., 1:561-573), followed by two C residues, the *Eco*RI recognition sequence GAATTC, and the rest of pUC 19.

1. pBI221 DNA was digested with *Eco*RI and *Bam*HI, and the 3506 bp fragment was separated from the 2163 bp small fragment by agarose gel electrophoresis, and then purified by standard methods. pRAJ275 (CLONTECH, Jefferson (1987), *supra*) DNA was digested with *Eco*RI and *Sal*I, and the 1862 bp fragment was purified from an agarose gel. These two fragments were mixed together, and complementary synthetic oligonucleotides having the sequence GATCCGGATCCG and TCGACGGATCCG were added. The fragments were ligated together and the ligation reaction was transformed into competent *E. coli* cells. A transformant harboring a plasmid having the appropriate

DNA structure was identified by restriction enzyme site mapping. This plasmid was named pKA881.

2. pKA881 DNA was digested with *Bal*I and *Eco*RI, and the 4148 bp large fragment was purified from an agarose gel. DNA of pBI221 was similarly digested, and the 1517 bp *Eco*RI/*Bal*I fragment was gel purified and ligated to the above pKA881 fragment, to generate plasmid pKA882.

3. pKA882 DNA was digested with *Sst*I, the protruding ends were made blunt by treatment with T4 polymerase, and the fragment was ligated to synthetic *Bam*HI linkers having the sequence CGGATCCG. An *E.coli* transformant that harbored a plasmid having *Bam*HI fragments of 3784 and 1885 bp was identified and named pKA882B.

4. pKA882 DNA was digested with *Pst*I, and the linear fragments were ligated to synthetic adaptors having the sequence CAGATCTGTGCA. An *E.coli* transformant was identified that harbored a plasmid that was not cleaved by *Pst*I, and that had a new, unique *Bgl*II site. This plasmid was named pKA882-Bg.

5. pKA882-Bg DNA was digested with *Eco*RI, and the linear fragments were ligated to synthetic adaptors having the sequence AATTGAGATCTC. An *E.coli* transformant was identified that harbored a plasmid that was not cleaved by *Eco*RI, and that generated *Bgl*II fragments of 3027 and 2658 bp. This plasmid was named pKA882-2xBg.

6. pKA882B DNA was digested with *Bam*HI and the mixture of fragments was ligated. An *E.coli* transformant harboring a plasmid that generated a single 3783 bp fragment upon digestion with *Bam*HI was identified and named p35S/*Nos*. This plasmid has the essential DNA structure of pBI221, except that the coding sequences of

the GUS gene have been deleted. Therefore, CaMV nucleotides 6605 t0 7439 are followed by the linker sequence GGGGACTCTAGAGGATCCCGAATTTCCCC, which is followed by the *Nos* Polyadenylation sequences and the rest of pBI221.

7. p35S/*Nos* DNA was digested with *Eco*RV and *Pst*I, and the 3037 bp fragment was purified and ligated to the 534 bp fragment obtained from digestion of p35S/En2 DNA (see Example 5, Section C.5) with *Eco*RV and *Pst*I. An *E.coli* transformant was identified that harbored a plasmid that generated fragments of 3031 and 534 bp upon digestion with *Eco*RV and *Pst*I, and the plasmid was named p35S En2/*Nos*. This plasmid contains the duplicated 35S promoter enhancer region described for p35S En2 in Example 5, Section C.5. The promoter sequences was separated from the *Nos* polyadenylation sequences by linker sequences that include a unique *Bam*HI site.

B. Plasmids utilizing the 35S promoter and the Agrobacterium ORF 25/26 Poly A sequences

The starting material is plasmid pIC 35. This plasmid contains the 845 bp *Sma*I/*Hind*III fragment from pUC 13 35S (-343) [see Example 12, Section C], ligated into the *Nru*I and *Hind*III sites of pIC 19R (Marsh, *et al.* (1984), Gene, 32:481-485), in the orientation such that the *Hind*III recognition site is maintained. The source of the *A. tumefaciens* ORF25/26 sequences is plasmid pIC1925. This plasmid contains the 713 bp *Hinc*II fragment comprising nucleotides 21728 to 22440 of *A. tumefaciens* pTi 15955 T-DNA (Barker *et al.*, Plant Molec. Biol., 2:335-350), ligated into the *Sma*I site of pIC 19H (Marsh, *et al.* (1984), *supra*), in the orientation such

that the *Bam*HI site of pIC 19H is adjacent to the ORF 25 end of the T-DNA fragment.

1. DNA of plasmid pIC 35 was digested with *Bam*HI, and ligated to a 738 bp fragment prepared by digestion of pIC1925 DNA with *Bam*HI and *Bgl*II. An *E.coli* transformant was identified that harbored a plasmid in which a *Bam*HI site was positioned between the 35S promoter fragment and the ORF 25/26 Poly A fragment. This plasmid was named pIC 19R35/A.

2. pIC 19R35/A DNA was digested with *Sma*I at its unique site, and the DNA was ligated to *Bgl*II linkers having the sequence CAGATCTG. The tandomization of these *Bgl*II linkers generates, besides *Bgl*II recognition sites, also *Pst*I recognition sites, CTGCAG. An *E.coli* transformant was identified that had at least two copies of the linkers (and new *Bgl*II and *Pst*I sites) at the position of the former *Sma*I site. This plasmid was named pIC35/A.

3. DNA of plasmid pIC 20R (Marsh, *et al.* (1984), Gene, 32:481-48514) was digested with *Nru*I and *Sma*I, and the blunt ends of the large fragment were ligated together. An *E.coli* transformant was identified that harbored a plasmid that lacked *Nru*I, *Sma*I, *Hin*dIII, *Sph*I, *Pst*I, *Sal*I, *Xba*I, and *Bam*HI sites. This plasmid was called pIC 20RD.

4. pIC 20RD DNA was digested with *Bgl*II, and was ligated to the 1625 bp *Bgl*II fragment of pIC35/A. An *E. coli* transformant was identified that harbored a plasmid that contained the 35S promoter/ORF 25 poly A sequences. Restriction enzym site mapping revealed these sequences to be in the orientation such that the unique *Kpn*I and

*Xho*I sites of pIC 20RD are positioned at the 3' end of the ORF 25 Poly A sequences. This plasmid was named pSG Bgl 3525 (Pst).

5. DNA of pSG *Bgl*II 3525 (Pst) was digested with *Bgl*II under conditions in which only one of the two *Bgl*II sites of the molecule were cleaved. The 4301 bp linear fragments were ligated to synthetic adapter oligonucleotides having the sequence GATCGTGATCAC, where the underlined bases represent the *Bcl*I recognition sequence. An *E.coli* transformant was identified that had a *Bcl*I site at the position of the former *Bgl*II site positioned 5' to the 35S promoter. This plasmid was named pSG 3525 a (*Pst*).

C. Construction of a doubly-enhanced CaMV 35S Promoter

The starting material is plasmid pUC13/35S (-343) as described by Odell *et al.* ((1985), Nature, 313:810-812). This plasmid comprises, starting at the 3' end of the *Sma*I site of pUC 13 (Messing, J. (1983) in "Methods in Enzymology" (Wu, R. *et al.*, Eds) 101:20-78), and reading on the strand contiguous to the noncoding strand of the Lac Z gene of pUC 13, nucleotides 6495 to 6972 of CaMV, followed by the linker sequence CATCGATG (which encodes a *Cla*I recognition site), followed by CaMV nucleotides 7089 to 7443, followed by the linker sequence CAAGCTTG, the latter sequence including the recognition sequence for *Hin*dIII, which is then followed by the remainder of the pUC 13 plasmd DNA.

1. pUC 13/35S (-343) DNA was digested with *Cla*I, and the protruding ends were made flush by treatment with T4 polymerase. The blunt-ended DNA was then ligated to

synthetic oligonucleotide linkers having the sequence CCCATGGG, which includes an *Nco*I r ognition site. An *E.coli* transformant was identified containing a plasmid (named pOO#1) having an *Nco*I site positioned at the former *Cla*I site.

2. pOO#1 DNA was digested with *Nco*I and the compatible ends of the large fragment were religated, resulting in the deletion of 70 bp from pOO#1, to generate plasmid pOO#1 *Nco*Δ.

3. pOO#1 *Nco*Δ DNA was digested with *Eco*RV, and the blunt ends were ligated to *Cla*I linkers having the sequence CATCGATG. An *E.coli* transformant harboring a plasmid having a new *Cla*I site at the position of the previous *Eco*RV site was identified, and the plasmid was named pOO#1 *Nco*ΔRV/*Cla*.

4. pOO#1 *Nco*Δ RV/*Cla* DNA was digested with *Cla*I and *Nco*I, and the small (268 bp) fragment was purified from an agarose gel. This fragment was then ligated to the 3429 bp *Cla*I/*Nco*I fragment of pUC 13/35S (-343) prepared by isolation from an agarose gel, and an *E.coli* transformant was identified that harbored a plasmid having *Cla*I/*Nco*I fragments 3429 and 268 bp. This plasmid was named pUC 13/35S En.

5. pUC 13/35S En DNA was digested with *Nco*I, and the protruding ends were made blunt by treatment with T4 polymerase. The treated DNA was then cut with *Sma*I, and was ligated to *Bgl*II linkers having the sequence CAGATCTG. An *E.coli* transformant was identified that harbored a plasmid in which the 416 bp *Sma*I/*Nco*I fragment had been replaced with at least two copies of the *Bgl*II link rs and named p35S En².

The DNA structure of p35S En² is as follows: eginning with the nucleotide that follows the third C residue of the *Sma*I site on the strand contiguous to the noncoding strand of the Lac Z gene of pUC 13; the linker sequence CAGATCTGCAGATCTGCATGGGCGATG, followed by a CaMV nucleotides 7090 to 7344, followed by a *Cla*I linker sequence CATCGATG, followed by CaMV nucleotides 7089 to 7443, followed by the *Hin*dIII linker sequence CAAGCTT, followed by the rest of pUC 13 sequence. This structure has the feature that the enhancer sequences of the CaMV 35S promoter, which lie in the region upstream of the *Eco*RV site in the viral genome (nucleotides 7090 to 7344), have been duplicated. This promoter construct incorporates the native 35S transcription start site, which lies 11 nucleotides upstream of the first A resdue of the *Hin*dIII site.

D. Construction of a synthetic untranslated leader

A DNA fragment was constructed that includes sequences which comprise the 5' untranslated leader portion of the major rightward transcript of the Maize Streak Virus (MSV) genome. The MSV genomic sequence was published by Mullineaux *et al.*, (1984), EMBO J., 3:3063-3068, and Howell (1984), Nucl. Acids Res., 12:7359-7375, and the transcript was described by Fenoll *et al.* (1988), EMBO J., 7:1589-1596. The entire sequence, comprising 154 bp, was constructed in three stages by assembling blocks (A, B, and C) of synthetic oligonucleotides.

1. The A Block: Complementary oligonucleotides having the sequence
GATCCAGCTGAAGGCTCGACAAGGCAGATCCACGGAGGAGCTGA
TATTTGGTGGACA and AGCTTGTCCACCAAATATCAGCTCCTCCGTGGATC
TGCCTTGTCGAGCCTTCAGCTG were synthesized and purified by

standard procedures. Annealing of th s nucleotides into double-stranded structures leaves 4-base sticky ends that are compatible with those generated by *Bam*HI on one end of the molecule (GATC), and with *Hin*dIII-generated single stranded ends on the other end of the molecule (AGCT) Such annealed molecules were ligated into plasmid pBluescript SK(-) [Stratagene Cloning Systems, La Jolla, CA], that had been digested with *Bam*HI and *Hin*dIII. An *E.coli* transformant harboring a plasmid containing the oligonucleotide sequence was identified by BamHI and *Hin*dIII restriction enzyme analysis, and the plasmid was named pMSV A.

2. The B Block: Complementary oligonucleotides having the sequences
AGCTGTGGATAGGAGCAACCCTATCCCTAATATACCAGCACCA
CCAAGTCAGGGCAATCCCGGG and TCGACCCGGGATTGCCCTGACTTGGTGG
TGCTGGTATATTAGGGATAGGGTTGCTCCTATCCAC were synthesized and purified by standard procedures. The underlined bases represent the recognition sequence for restriction enzymes *Sma*I and *Xma*I. Annealing of these nucleotides into double-stranded structures leaves 4-base sticky ends that are compatible with those generated by *Hin*dIII on one end of the molecule (AGCT), and with *Sal*I-generated sticky ends on the other end of the molecule (TCGA).

DNA of pMSV A was digested with *Hin*dIII and *Sal*I, and was ligated to the above annealed oligonucleotides. An *E.coli* transformant harboring a plasmid containing the new oligonucleotides was identified by restriction enzyme site mapping, and was named pMSV AB.

3. The C Block: Complementary oligonucleotides having the sequences CCGGGCCATTTGTTCCAGGCACGGGATAAGCA TTCAGCCATGGG ATATCAAGCTTGGATCCC and TCGAGGGATCCAAGCTTGATATCCCATGGC TGAATGCTTATCCCGTGCCTGGAACAAATGGC were synthesized and purified by standard procedures. These oligonucleotides incorporate bases that comprise recognition sites (underlined) for *Nco*I (CCATGG), *Eco*RV (GATATC), *Hin*dIII (AAGCTT), and *Bam*HI (GGATCC). Anealing of these nucleotides into double-stranded structures leaves 4-base sticky ends that are compatible with those generated by *Xma*I on one end of the molecule (CCGG), and with *Xho*I-generated sticky ends on the other end of the molecule (TCGA). Such annealed molecules were ligated into pMSV AB DNA that had been digested with *Xma*I and *Xho*I. An *E.coli* transformant harboring a plasmid containing the oligonucleotide sequence was identified by restriction enzyme site analysis, and DNA structure was verified by sequence analysis. The plasmid was named pMSV CPL; it contains the A, B and C blocks of nucleoides in sequential order ABC. Together, these comprise the 5' untranslated leader sequence ("L") of the MSV coat protein ("CP") gene. These correspond to nucleotides 167 to 186, and nucleotides 188 to 317 of the MSV sequence of Mullineaux *et al.*, (1984), *supra*, and are flanked on the 5' end by the *Bam*HI linker sequence GGATCCAG, and on the 3' end by the linker sequence GATATCAAGCTTGGATCCC. An A residue corresponding to base 187 of the wild type MSV sequence was inadvertently deleted during cloning.

4. *Bgl*II Site Insertion

pMSV CPL DNA was digested at the *Sma*I site corresponding to base 277 of the MSV genomic sequence

(Mullineaux, *et al.* (1984), *supra*), and the DNA was ligated to *Bgl*II linkers having the sequence CAGATCTG. An *E. coli* transformant harboring a plasmid having a unique *Bgl*II site at the position of the former *Sma*I site was identified and verified by DNA sequence analysis, and the plasmid was named pCPL-Bgl.

E. Construction of a deleted version of the maize alcohol dehydrogenase 1 (Adh1) intron 1

The starting material is plasmid pVW119. This plasmid contains the DNA sequence of the maize Adh 1.S gene intron 1 from nucleotides 119 to 672, and was described in Callis *et al.* (1987), Genes and Devel., 1:1183-1200. The sequence following base 672 of Dennis *et al.* ((1984), Nucl. Acids Res., 12:3983-4000) is GACGGATCC, where the underlined bases represent a *Bam*HI recognition site. The entire intron 1 sequence, including 14 bp of exon 1, and 9 bp of exon 2, was obtained from this plasmid on a 556 bp fragment following digestion with *Bcl*I and *Bam*HI.

1. Plasmid pSG 3525 a (Pst) DNA (see Example, section B.5) was digested with *Bam*HI and *Bcl*I, and the 3430 bp fragment was purified from an agarose gel. pVW119 DNA was digested with *Bam*HI and *Bcl*I, and the gel purified fragment of 556 bp was ligated to the above 3430 bp fragment. An *E. coli* transformant was identified that harbored a plasmid that generated fragments of 3430 and 556 bp upon digestion with *Bam*HI and *Bcl*I. This plasmid was named pSG Adh A1.

2. pSG Adh A1 DNA was digested with *Hin*dIII, [which cuts between bases 209 and 210 of the Dennis *et al.* ((1984), *supra*) sequence, bottom strand], and with *Stu*I, which cuts between bases 554 and 555. The ends were

made flush by T4 polymerase treatment, and then ligated. An *E.coli* transformant harboring a plasmid lacking *Hin*dIII and *Stu*I sites was identified, and the DNA structure was verified by sequence analysis. The plasmid was named pSG Adh A1D. In this construct, 344 bp of DNA have been deleted from the interior of the intron 1. The functional intron sequence is obtained on a 213 bp fragment following digestion with *Bcl*I and *Bam*HI.

3. pCPL-Bgl DNA (see Example 5, Section D.4), was digested with *Bgl*II, and the linearized DNA was ligated to the 213 bp *Bcl*I/*Bam*H fragment containing the deleted version of the Adh 1.S intron 1 sequences from pSG Adh A1D. An *E.coli* transformant was identified by restriction enzyme site mapping that harbored a plasmid containing the intron sequences ligated into the *Bgl*II site, in the orientation such that the *Bgl*II/*Bcl*I juncture was nearest the 5' end of the MSV CPL leader sequence, and the *Bgl*II/ *Bam*HI juncture was nearest the 3' end of the CPL. This orientation was confirmed by DNA sequence analysis. The plasmid was named pCPL AlIlD. The MSV leader/intron sequences is obtained from this plasmid by digestion with *Bam*HI and *Nco*I, and purification of the 373 bp fragment.

F. Construction of plant expression vectors based on the enhanced 35S promoter, the MSV CPL, and the deleted version of the Adh 1 intron 1

1. DNA of plasmid p35S En2/Nos (see Example 5, Section A.7) was digested with *Bam*HI, and the 3562 bp linear fragment was ligated to a 171 bp fragment prepared from pMSV CPL DNA digested with *Bam*HI. This fragment contains the entire MSV CPL sequence described in

Section D.3. An *E.coli* transformant was identified by restriction enzyme site mapping that harbored a plasmid that contained these sequences in an orientation such that the *Nco*I site was positioned near the Nos Poly A sequences. This plasmid was named p35S En2 CPL/Nos. It contains the enhanced version of the 35S promoter directly contiguous to the MSV leader sequences, such that the derived transcript will include the MSV sequences in its 5' untranslated portion.

2. DNA of plasmid pKA882 (see Example 5, section A.2) was digested with *Hin*dIII and *Nco*I, and the large 4778 bp fragment was ligated to an 802 bp *Hin*dIII/*Nco*I fragment containing the enhanced 35S promoter sequences and MSV leader sequences from p35S En2 CPL/Nos. An *E. coli* transformant harboring a plasmid that contained fragments of 4778 and 802 bp following digestion with *Hin*dIII and *Nco*I was identified, and named pDAB 310. In this plasmid, the enhanced version of the 35S promoter is used to control expression of the GUS gene. The 5' untranslated leader portion of the transcript contains the leader sequence of the MSV coat protein gene.

3. DNA of plasmid pDAB 310 was digested with *Nco*I and *Sst*I. The large 3717 bp fragment was purified from an agarose gel and ligated to complementary synthetic oligonucleotides having the sequences CGGTACCTCGAGTTAAC and CATGGTTAACTCGAGGTACCGAGCT. These oligonucleotides, when annealed into double stranded structures, generate molecules having sticky ends compatible with those left by *Sst*I (AGCT), on one end of the molecule, and with *Nco*I (CATG) on the other end of the molecule. An *E.coli* transformant was identified that harbored a plasmid containing sites for enzymes *Sst*I (AGCT), *Nco*I (CATG), *Kpn*I (GGTACC), *Xho*I (CTCGAG), and *Hpa*I (GTTAAC), and

the DNA structure was verified by sequence analysis. This plasmid was named pDAB 1148.

4. DNA of plasmid pDAB 1148 was digested with *Bam*HI and *Nco*I, the large 3577 bp fragment was purified from an agarose gel and ligated to a 373 bp fragment purified from pCPL AII1D (see Example 5, Section E.3) following digestion with *Bam*HI and *Nco*I. An *E.coli* transformant was identified that harbored a plasmid that generated fragments of 3577 and 373 bp following digestion with *Bam*HI and *Nco*I, and the plasmid was named pDAB 303. This plasmid has the following DNA structure: beginning with the base after final G residue of the *Pst*I site of pUC 19 (base 435), and reading on the strand contiguous to the coding strand of the Lac Z gene, the linker sequence ATCTGCATGGGTG, nucleotides 7093 to 7344 of CaMV DNA, the linker seqeunce CATCGATG, nucleotides 7093 to 7439 of CaMV, the linker sequence GGGGACTCTAGAGGATCCAG, nucleotides 167 to 186 of MSV, nucleotides 188 to 277 of MSV, a C residue followed by nucleotides 269 to 359 of Adh 1S intron 1, nucleotides 704 to 821 of maize Adh S intron 1, the linker sequence GACGGATCTG, nucleotides 278 to 317 of MSV, the linker sequence GTTAACTCG AGGTACCGAGCTCGAATTTCCCC, nucleotides 1298 to 1554 of Nos, and a G residue followed by the rest of the pUC 19 sequence (including the *Eco*RI site). The unique *Eco*RI and *Bgl*II sites of pDAB303 were converted to *Not*I sites, respectively, by oligonucleotides containing the recognition sequence GCGGCCGC to form pDAB303-Not.

G. Construction of plant transformation vectors containing the bar gene of *Streptomyces hygroscopicus*

The starting material is plasmid pIJ4104 (White, *et al.* (1990), Nucl. Acids Res., 18:1062), which

contains the coding region of the *bar* gene of *S. hygroscopicus*, which encodes the enzyme phosphinothricin acetyl transferase (PAT).

1. DNA of plasmid pIJ4104 was digested with *Sma*I, and the 569 bp fragment was purified from an agarose gel. DNA of plasmid pSG 3525 a⁻(Pst) (see Example 5, Section B.5) was linearized by digestion at the unique *Hinc*II that lies between the 35S promoter and ORF 25 poly A sequences, and the linear fragment was ligated to the 569 bp bar gene fragment. An *E.coli* transformant was identified by restriction enzyme site mapping that harbored a plasmid containing the bar gene in the orientation such that *Bgl*II digestion generated fragments of 4118 and 764 bp. This plasmid was named pDAB 218.

2. DNA of plasmid pDAB 218 was digested with *Bcl*I, and the linear fragment of 4882 bp was ligated to a 3133 bp *Bgl*II fragment prepared from DNA of pKA882-2xBg (see Example 5, Section A.5). The latter fragment contains the GUS coding region, under the transcriptional control of the 35S promoter, with the Nos Poly A transcription termination signals. An *E.coli* transformant was identified that contained the GUS and PAT coding regions, and restriction enzyme recognition site mapping revealed that both coding regions were encoded by the same DNA strand. This plasmid was named pDAB 219.

3. DNA of plasmid pDAB 219 was used as the template for the polymerase chain reaction (Saiki *et al.*, (1988), Science, 239:487-491) using as primers the synthetic oligonucleotides: i) CTCGAGATCTAGATATCGATGAATTCCC, and ii) TATGGATCCTGTGATAACCGACATATGCCCCGGTTTCGTTG. Primer i) represents nucleotides 419 to 446 of pDAB 219, and

includes bases corresponding to the recognition sites of *Xho*I (CTCGAG), *Bgl*II (AGATCT), *Xba*I (TCTAGA), *Eco*RV (GATATC), *Cla*I (ATCGAT), and *Eco*RI (GAATTC). The single underlined bases in primer ii) represent the recognition sequence of *Bam*HI, and the double underlined bases represent nucleotides 1138 to 1159 of pDAB 219, and correspond to nucleotides 21728 to 21749 of the ORF 25 Poly A fragment (see Example 5, Section B). PCR amplification generated a product of 760 bp.

4. DNA of plasmid pDAB 219 was digested with *Bgl*II, the 7252 bp fragment was purified from an agarose gel, and ligated to the 747 bp fragment generated by digestion of the above PCR product by *Bgl*II and *Bam*HI. An *E.coli* transformant was identified that harbored a plasmid containing a unique *Bgl*II site positioned at the 3' end of the ORF 25 Poly A fragment. The DNA structure of the 3' end of the PAT coding sequence was confirmed by DNA sequence analysis. This plasmid was named pDAB 219Δ.

The DNA sequence of pDAB 219Δ is as follows: Beginning with the base following the last A residue of the *Xba*I site on the Lac Z coding strand of pIC 20R (Marsh, *et al.* (1984), Gene, 32:481-485), the linker TCCTGATCTGTGCAGGTCCCC, followed by CaMV nucleotides 6605 to 7439, followed by the linker sequence GGGGACTCTAGAGGATCCGGATCCGTCGACATGGTC, followed by the rest of the coding region of GUS with 44 bp of 3' flanking *E.coli* genomic DNA (nucleotides 306 to 2152 of Jefferson *et al.* (1986), (Proc. Natl. Acad. Sci., 83:8447-8451). The underlined bases represent the codons for the first two amino acids of th GUS protein, the second of which was changed from leucine in the original *E.coli* uidA gene (Jefferson *et al.* (1986), *supra*) to valine in

pRAJ275 (Jefferson, (1987), *supra*). These bases are followed by the linker sequence GGGGAATTGGAGAGCTCGA ATTTCCCC, then by bases 1298 to 1554 of the Nos Poly A sequence (DePicker, *et al.* (1982), J. Molec. Appl. Genet., 1:561-5736). The linker sequence GGGAATTGAGATCAGGATCTCGAGCTCGGG is followed by bases 495 to 6972 of CaMV, the linker CATCGATG, and CaMV bases 7090 to 7443. These bases are followed by the linker CAAGCTTGGCTGCAGGTC, then by bases corresponding to nucleotides 20 to 579 of the *bar* clone in pIJ4104 (White, *et al.* (199)), Nucl. Acids Res., 18:1062), the linker CTGTGATAACC, ORF 25/26 poly A nucleotides 21728 to 22440 (Barker, *et al.* (1983), Plant Molec. Biol., 2:335-3501), the linker GGGAATTCATCGATATCTAGATCTCGAGCTCGGGGTACCGAGCTCGAATTC, and the rest of pIC 20R. The *Bgl*II recognition site (underlined) represents a unique site into which other genes may be introduced. A partial restriction map of pDAB 219Δ is appended. An oligonucleotide containing the recognition sequence for the restriction enzyme *Not*I (GCGGCCGC) was introduced into the *Bgl*II site to give the plasmid pDAB219Δ-Not (shown in Figure 7).

Example 6: Construction of a cDNA library from potato tuber skin tissue

A. RNA Purification

The skin and outer cortex tissue was harvested from 4 cm potato tubers (*Solanum tuberosum* cv. Superior) and immediately frozen in liquid nitrogen. Frozen tissue was ground in a mortar to a fine powder under liquid nitrogen. Five grams of tissue were extracted with a volume of 50 mM Tris-HCl pH 8.0, 4% para-amino salicylic acid, 1% Tri-isopropylnaphthalenesulfonic

acid, 10 mM dithrothreitol, and 10 mM sodium metabisuffite. Homogenate was then extracted with an equal volume of phenol containing 0.1% 8-hydroxyquinoline. After centrifugation, the aqueous layer was extracted with an equal volume of phenol containing chloroform:isoamy alcohol (24:1), followed by extraction with chloroform:octanol (24:1). Subsequently, 7.5 M ammonium acetate was added to a final concentration of 2.5 M. The RNA was precipitated overnight at -20°C, collected by centrifugation, reprecipitated with 2.5 M ammonium acetate and washed with 70% ethanol. The dried RNA was resuspended in water and stored at -80°C. Poly A^+ RNA was isolated using Hybond mAP™ messenger affinity paper (Amersham).

B. Library Construction and Screening

cDNA was synthesized using 5 µg of Poly A^+ RNA and the ZAP-cDNA™ synthesis kit (Stratagene). Size selected cDNA was ligated to 2 µg of UniZap XR™ vector arms (Stratagene), and packaged into phage particles with Gigapack Gold™ packaging extract (Stratagene). About 4.2×10^6 putative clones were obtained after packaging. The plate amplified library contained approximately 5.0×10^{10} pfu/ml when titered using *E. coli* PLK-F'™ cells (Stratagene) as the host strain. About 1.5×10^5 plaques were screened (1.5×10^4 per 10 cm filter) using the picoBlue™ immunoscreening kit (Stratagene) and polyclonal rabbit antisera against the 87 kD PPI protein (Berkeley Antibody Company). Plaques with the strongest reaction were rescreened an additional three times to ensure plaque purity. DNA was subcloned from these plaques as a plasmid, by the

excision mechanism engine red into the UniZap lambda vector.

C. Plasmid Isolation and Confirmation by Southern Analysis

Plasmid DNA was isolated by the alkaline lysis method (Sambrook *et al.* (1989), *supra*). Insert size was determined by complete digestion with *Eco*RI/*Xho*I (New England Biolabs) and agarose gel electrophoresis. Nine of the largest cDNAs including one negative control were digested as above, electrophoresed through 1% agarose, and blotted to nitrocellulose (Schleicher & Schuell). Two oligonucleotide probes based on C-terminal and N-terminal regions respectively, of the amino acid sequence of PPI were synthesized on the DNA synthesizer (Applied Biosystems Inc. Model 28A) and were labeled at their 5' ends with (^{32}P) ATP using T_4 polynucleotide kinase. Filters were prehybridized individually at 37°C for 2 hours in 50 mls of 6x SSC (0.9 M Sodium Chloride, 0.09 M Sodium Citrate), 5x Denhardts solution, 20 mM $NaPO_4$ pH 7.4, 0.5% SDS and 250 µg/ml salmon sperm DNA. Hybridization was performed in the same solution at 37°C overnite with the addition of 12 x 10^6 cpm of each probe to the individual boxes. The filters were then washed 3x at room temperature in 2x SSC (0.45 M Sodium Chloride, 0.045 M Sodium Citrate)+0.1% SDS, then 1x at 37°C for 30 minutes in 2x SSC + 0.1% SDS. They were exposed to Kodak X-OMAT film for 24 hours.

Two of the cDNAs showed strong homology to both the C-terminal and N-terminal probes. The longest of these clones, was determined to be 1 kb in length, and

was named potato cystatin 1 (PC1) this clone was further characterized by DNA sequencing.

Example 7: A gene coding for a single cystatin-like unit of PPI

The complete sequence of the insert in PC1 was determined, using Erase-A-Base™ kit (Promega, Madison, WI) to generate nested deletions, and Sequenase™ kit (US Biochemical, Cleveland, OH). This DNA was derived from the 3' end of a PPI message. Its 5' end begins within the region coding for a cystatin-like unit, and is followed by sequence coding for two complete units at the carboxy-terminus. It also includes the 3' non-translated region and a poly A tail. The penultimate unit was PCR amplified (Perkin-Elmer, Norwalk, CT), using 25 cycles of 1 minute at 94°C, 1 minute at 50°C and 3 minutes at 72°C. The primers (GTCATAGAATTCAACGCCATGGCTGGTGATGTCCCAATACTC and GGAACAATGATAAGAGCTCATTACTTTGCACTATCATCACC) were assembled on a 380A DNA Synthesizer (Applied Biosystems, Foster City, CA). They add an in-frame initiator ATG within a *Nco*I site at the 5' end, and a TAA stop codon and a *Sst*I site at the 3' end. The amplified DNA was cloned into pCR1000 (Invitrogen, San Diego, CA). The region coding for a complete cystatin-like unit was transferred on a *Nco*I-*Sst*I fragment into the plant expression vector, pDAB303-Not. The resultant plasmid, pDAB1189, was used to demonstrate the functional integrity of the constructed gene. When introduced into protoplasts of cultured corn cells, it directed synthesis of a polypeptide which cross-reacted specifically with anti-PPI antiserum.

Example 8: A gene coding for a secreted cystatin-like unit of PPI.

A sequence coding for a 23 amino acid amino-terminal signal peptide was added to the PPI-coding sequence in pDAB1189. The signal sequence, derived from a prepatatin cDNA clone (pDAB1008), was assembled from synthetic oligonucleotides in the following three steps:

1. A vector lacking *Afl*III sites was prepared from pKK233-2 (Pharmacia, Piscataway, NJ) by cutting the DNA with *Afl*III, filling in the ends with Klenow fragments of DNA polymerase I, and recircularizing the plasmid with T4 DNA ligase. Enzymes were used according to the manufacturer's protocols.

2. Two oligonucleotides (CATGGCAACTACTAAATCTTTTTTAATTTTATTTTTTATGATATTAGCAACTACT AGTTCAACATGTTAACGGTACCCGGGCCATGGA and AGCTTCCATGGCCCGGGTACCGTTAACATGTTGAACTAGTAGTTGCTAATATCAAA AAATAAAATTAAAAAAGATTTAG) were cloned into the pKK233-2 derivative, between the *Nco*I and *Hin*dIII sites. The resultant plasmid, pDAB1079, contains the first 22 amino acids of the signal between a *Nco*I and *Afl*III site (underlined).

3. The signal sequence was completed by cloning the oligonucleotides CATGTGCCATGG and AGCTTCCATGGCA between the *Afl*III and *Hin*dIII sites of pDAB1079. They add the last codon, followed by an *Nco*I site containing an in-frame ATG codon. The complete signal, on an *Nco*I fragment, was introduced from the PPI-coding sequence in pDAB1189 to generate pDAB1209. The precursor cystatin-like unit encoded by this plasmid has the following amino-terminal sequence:

pDAB1209: MATTKSFLILFFMILATTSSTCAMAGDVPILGG...
cf.pDAB1089: MAGDVPILGG...

Example 9: A gene with introns coding for full-length PPI.

The tuber cDNA library was re-screened with the 650 bp *Eco*RI fragment of PC1 to obtain longer cDNA clones. The library was plated on *E.coli* PLK-F' cell, as described in Stratagene's Uni-Zap Manual. Only a low density, of about 3000 phage per plate (80 mm diameter), was used because a high proportion of hybridizing plaques was anticipated. The plaques were transferred to Nytran™ filters (Schleicher Schuell, Keene, NH) and fixed by UV irradiation in a Stratalinker™ apparatus (Stratagene). The gel-purified PC1 fragment was labeled with digoxygenin by Klenow polymerization from random primers (Boehringer-Mannheim, Indianapolis, IN). Filters were hybridized according to the manufacturer's protocols, with stringent washes in 2xSSC, 0.1% SDS at 65°C. Binding was detected with anti-digoxygenin antibody coupled to alkaline phosphatase (Boehringer-Manneheim)

Hybridizing cDNAs were excised as phagemids, by co-infecting with the phage and a helper phage (R408), as described in the Stratagene Uni-Zap™ Manual. An extra step of retransformation at low density was necessary to eliminate helper from the phagemids. The sizes of the cDNA inserts were determined by digestion with *Nco*I and *Xho*I, followed by agaraose gel electrophoresis. The longest insert (in pDAB1034) coded for almost 4 cystatin-like units, the non-translated 3' r gion and a poly A tail.

A *Bsm*I fragment of pDAB1034, coding for 3 cystatin-lik units, was gel-purified and labeled with digoxygenin, as above. This probe was used to screen a library of potato genomic DNA fragments in phage lambda Fix™ II (Stratagene). The same conditions as above were used, except that phage were plated at a density of 20,000 per plate. One clone, BZ2, was purified by 3 cycles of replating at low density, and hybridization. A large-scale preparation of this phage was made by infecting exponentially-growing *E.coli* LE392 cells with 10 phage/cell. To maximize yield, the culture was diluted 10-fold after 2.5 hours growth at 250 rpm and 37°C, and re-incubated overnight. Cell lysis was completed by the addition of chloroform (4 mL/L). Debris was removed by low-speed centrifugation after treatment with 6% NaCl for 1.5h at room temperature. The phage were precipitated by 5 hour treatment on ice with 7% PEG6000 (Fluka AG, Buchs, Switzerland), then collected by 10 minutes centrifugation at 12,000 x g. Phage were resuspended in 100mM NaCl, 8mM MgSO4, 0.01% gelatin, 50mM Tris.HCl pH 7.5, and purified by equilibrium density centrifugation in CsCl (43% w/w) at 155,000 x g overnight. DNA was prepared from the phage in buoyant density CsCl by ethanol precipitation, after overnight incubation at room temperature in 50% formamide, 10 mM EDTA, 0.1M Tris.HCl, pH 8.5.

BZ2 DNA was characterized by standard restriction enzyme analysis and Southern blotting (Ausubel *et al.*, 1987; Sambrook *et al.* (1989), *supra*). The phage contained a 13 kb insert, with the cystatin-like homology localized in a 5 kb central region. Fragments spanning this hybridizing region were subcloned into pUC18 or pBluescript™ II plasmids (Stratagene), and

sequenced as described above. The PPI gene in BZ2 consists of 9 exons and 8 intr ns. It codes for a polypeptide consisting of 8 cystatin-like units, with no other domains. The DNA for each unit is interrupted by an intron, in the same position as an intron found in the oryzacystatin genes (Kondo (1991), FEBS Letts 278:87-90; and Kondo (1899), Gene, 81:259-265).

The 5' end of the coding sequence was modified to form a *Nco*I site around the initiator ATG, by PCR amplification using the primers TTCGCAGCCATGGCAATCG and TTTAAACTCCAAACTAGAATC. The amplified fragment was cloned into pTA1, a pUC19 derivative, such that the 3' end of the cloning sequence was adjacent to the *Bam*HI site in the polylinker. The region between the unique *Hpa*I site in the insert and the *Bam*HI site in the vector was replaced with a 5 kb *Hpa*I-*Bam*HI fragment from pDAB1286, which contained the 6.2 kb *Xho*I fragment of BZ2 in pBluescript II. This reconstituted the entire PP1 gene, along with 1.5 kb of 3' non-translated sequence. The mutated coding region was then transferred on a *Nco*I fragment to the plant expression vector, pDAB303-Not. The expression cassette from this plasmid was transferred as a *Not*I fragment into the vector pDAB219Δ-Not. The resultant plasmid, pDAB1400, was used for transformation of the maize plants.

Example 10: A gene without introns for full-length PPI.

The introns is removed from a cloned PPI genomic sequence by site-directed mutagenesis, using primers to loop out the DNA between coding regions. Such changes are made with a Transformer Mutagenesis Kit (Clontech, Palo Alto, CA). A plasmid containing the PPI gene with introns is mutagenized by *in vitro* replication

from two primers. One primer (GGATCCCCGGATACCGAGCTC) eliminates the unique *Kpn*I site in the vector. The second primer (GATTATAATAAGAAAGAGAATGCTCATTTGGAGTTT) is homologous to the coding sequences flanking intron 6 of the PPI gene. Mutant plasmids lacking intron 6 is selected by resistance to *Kpn*I digestion. Successive rounds of mutagenesis may be used to remove all the introns from the PPI genomic clone. The coding sequence may then be transferred into pDAB303-Not for expression in plant cells. The expression cassette from this plasmid was transferred as a *Not*I fragment into the vector pDAB219Δ-Not. The resultant plasmid was used for transformation of the maize plants.

Example 11: Development of a transgenic maize plant expressing a potato papain inhibitor

A. Establishment of Friable, Embryogenic Callus Cultures

Friable, embryogenic maize callus are initiated from immature embryos of the genotype B73 x A188. Seed of the dent-corn inbred, B73, and the sweet-corn inbred, A188, are obtained from Holden's Foundation Seeds, Inc., Williamsburg, IA and the University of Minnesota, Crop Improvement Association, St. Paul, MN, respectively. Seed are sown individually in pots containing approximately 18 kg of dry soil mix (Conrad Fafard, Inc., Springfield, MA) moistened and adjusted to pH 6.0. The plants are maintained in a greenhouse under a 16/8 photoperiod. Ambient daylight is supplemented with a combination of high pressure sodium and metal halide lamps such that the minimum light intensity 2 m above pot level is 1,500 ft-candles. Greenhouse temperature is maintained within 3°C of 38°C during the day and 22°C

at night. The plants are irrigated as needed with a solution containing 400 mg/L of 20-20-20 fertilizer (W.R. Grace & Co., Fogelsville, PA) plus 8 mg/L chelated iron (Ciba-Geigy, Greensboro, NC).

Approximately 50-60 days after planting, male inflorescences (tassels) are shedding pollen and silks have emerged from female inflorescences (ears). Pollen is collected by placing a paper bag over the tassel of a plant of the inbred line A188. A female plant of the inbred line B73 is prepared for pollination on the day before pollen availability by cutting off the tip of the husks and silks of an unfertilized ear shoot. The next day, after the silks have grown to form a thick "brush" all the same length, pollen is carefully applied to the silks and the entire ear is covered with a paper bag.

When the developing hybrid embryos reach a length of approximately 1.5-2.0 mm (10-14 days after pollination), the ear is excised and surface sterilized by emersion in 70% v/v ethanol for 10 minutes followed by soaking in 20% v/v commercial bleach (1% sodium hypochlorite) for 30 minutes. Following a sterile, distilled water rinse, immature embryos are aseptically isolated and placed onto a "callus" medium with the embryo axis in contact with the medium (scutellar-side away from the medium). The "callus" medium consists of the following components: N6 basal salts and vitamins (Chu et al., (1978) Proc. Symp. Plant Tissue Cult., Science Press, Peking, pp 43-56) 20 g/L sucrose, 691 mg/L proline, 100 mg/L casein hydrolysate, 1 mg/L 2,4-dichloro-phenoxyacetic acid (2,4-D), and 2.5 g/L gelrite (Kelco, Inc., San Diego, CA) adjusted to pH 5.8.

The immature, hybrid embryos are incubated at 28°C in the dark for 10-30 days during which time callus tissue, displaying various types of morphology, proliferates from the scutellar region. The callus tissue produced during this time is classified into three distinct types: i) soft, granular, translucent callus lacking any apparent morphological organization (known as non-embryogenic); ii) compact, nodular, yellowish-to-white callus consisting of groups of somatic embryos (often fused) with distinct scutellar- and coleoptile-like structures (known as Type I); and iii) soft callus with numerous globular and elongated somatic embryos on suspensor-like structures (known as Type II). Type II callus is the most suitable for establishing friable, embryogenic cultures. Sometimes entire scutella will proliferate with this type of tissue or at times only small sectors exhibiting this morphology will develop. At this point, selective sub-culture is necessary whereby only tissue with well-defined globular and elongated somatic embryos along with some subtending undifferentiated, soft tissue is transferred to fresh "callus" medium.

Every 10-14 days, the callus is sub-cultured to fresh "callus" medium being careful to select only tissue of the correct morphology. For the first 8-10 weeks, selection is for Type II callus only, to increase the amount of tissue and to select against non-embryogenic and Type I. At each sub-culture, less than 100 mg of tissue is typically selected from callus that has reached a size of 1 g. Thus, the amount of Type II callus will not increase to more than 1 g for the first 8-12 weeks due to the strict selection for tissue type. During the first 3 months, some lines (a line is defined

as originating from a single hybrid embryo) will be discarded if they lose their Typ II morphology. At about 8-16 weeks in well established Type II cultures, selection of a different type of tissue can proceed. This tissue (known as Type III) is different from Type II in that it is somewhat more homogeneous in morphology and relatively undifferentiated with no visible somatic embryos. The color will vary from light-to-bright yellow. Normally, it takes about 16-20 weeks to get this homogeneous, Type III tissue in sufficient amounts for routine experimentation (0.5-1.0 g).

During the 14-20 week period of Type III callus establishment, more lines are discarded if they revert to Type II or Type I after repeated selection. At 14-20 weeks of age, the cultures is checked for their ability to regenerate plants (see Example 11, Section C). Lines that do not regenerate are discarded. Cultures capable of maintaining Type III morphology and regenerating plants are referred to as friable, embryogenic callus.

B. Transformation via Microparticle Propulsion

Plasmid DNA is adsorbed onto the surface of gold particles prior to use in transformation experiments. The gold particles are spherical with diameters ranging from about 1.5-3.0 microns in diameter (Aldrich Chemical Co., Milwaukee, WI). Adsorption is accomplished by adding 74 uL of 2.5 M calcium chloride and 30 uL of 0.1 M spermidine to 300 uL of DNA/gold suspension (70 ug pDAB1400, 0.01 M Tris buffer, and 1mM EDTA). The DNA-coated gold particles are vortexed immediately, then allowed to settle to the bottom of an Eppendorf tube and the resultant clear liquid is completely drawn off. The DNA-coated gold particles are

then resuspended in 1 mL of 100% ethanol. The suspension is then diluted to 15 mg DNA/gold per mL of ethanol for use in microparticle propulsion experiments.

Approximately 250 mg of friable, embryogenic callus tissue, 5-7 days following sub-culture, is arranged in a thin layer on a 1 cm diameter piece of filter paper (Schleicher and Schuell, Inc., Keene, NH) placed on the surface of "callus" medium. The callus tissue is allowed to dry out slightly by allowing the plates to stand uncovered in a laminar flow hood for several minutes before use. In preparation for particle bombardment, the callus is covered with a 104 micron stainless steel screen. The DNA-coated gold particles are accelerated at the friable, embryogenic callus tissue using the particle bombardment apparatus described in European Patent Application EP 0 405 696 A1. Each callus tissue sample is bombarded 10-15 times with each bombardment delivering approximately 1 uL of DNA-coated gold suspension.

III. Selection of Transformed Tissue and Plant Regeneration

After bombarding the sample, callus tissue is allowed to incubate for 1-2 days under the conditions described previously (see Example 11, Section I). After 1-2 days, each tissue sample is divided into approximately 60 equal pieces (1-3 mm diameter) and transferred to fresh "callus" medium containing 30 mg/L Basta. Every three weeks, callus tissue is non-selectively transferred to fresh Basta-containing "callus" medium. At this concentration of herbicide, very little growth is observed. After 8-16 weeks, sectors proliferating from a background of growth

inhibited tissue is observed. This tissue is isolated from the other callus and maintained separately on Basta-containing "callus" medium and selectively sub-cultured every 10-14 days. At this point, a histochemical assay for gus expression is performed by placing small samples of callus tissue into 24-well microliter dishes (Corning, New York, NY) containing approximately 500 uL of assay buffer (0.2 M sodium phosphate pH 8.0, 0.1 mM each of potassium ferricyanide and potassium ferrocyanide, 1.0 M sodium EDTA, and 1 mg/L 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide). Potato papain inhibitor gene expression is also assayed via immunoblot analysis with potato papain inhibitor antiserum.

Basta-resistant, gus- and potato papain inhibitor-positive callus is selectively sub-cultured to "induction" medium and incubated at 28°C in low light (125 ft-candles) for one week followed by one week in high light (325 ft-candles) provided by cool fluorescent lamps. The "induction" medium is composed of MS salts and vitamins (Murashige and Skoog, 1962), 30 g/L sucrose, 100 mg/L myo-inositol, 5 mg/L benzyl-amino purine, 0.025 mg/L 2,4-D, 2.5 g/L Gelrite adjusted to pH 5.7. Following this two week induction period, the callus is then non-selectively transferred to "regeneration" medium and incubated in high light at 28°C. The "regeneration" medium is composed of MS salts and vitamins, 30 g/L sucrose, and 2.5 g/L gelrite adjusted to pH 5.7. Every 14-21 days the callus is subcultured to fresh "regeneration" medium selecting for tissue which appears to be differentiating leaves and roots. Both "induction" and "regeneration" media contain 30 mg/L Basta. Plantlets are transferred to 10

cm pots containing approximately 1 kg of dry soil mix, moistened thoroughly, and covered with clear plastic cups for approximately 4 days. At the 3-5 leaf-stage, plants are transplanted to larger pots and grown to maturity as previously described (see Example 11, Section A). Self- or sibling-pollinations is performed on plants regenerated from the same culture. Crosses to non-transformed parental lines (i.e., B73 or A188) can also be performed in order to obtain transgenic progeny analysis.

IV. Confirmation of Potato Papain Inhibitor Gene Integration

To confirm the presence of the potato papain inhibitor gene in regenerated plants and progeny, Southern blot analysis of genomic DNA is performed. DNA for each plant is prepared from lypholized leaf tissue as follows. Approximately 500 mg of tissue is placed into a 16 mL polypropylene tube (Becton Dickenson, Lincoln Park, NJ) into which is added 9 mL of CTAB extraction buffer (6.57 mL water, 0.9 mL of 1.0 M Tris pH7.5, 1.26 mL of 5 M sodium chloride, 0.18 mL of 0.5 M EDTA, 0.09 g mixed alkyl tri-methyl ammonium bromide, and 0.09 mL beta-mercaptoethanol) and immediately incubated in a 60°C water bath with occasional mixing. After about 60 minutes, 4.5 mL of 24:1 chloroform/octanol is added and gently mixed for approximately 5 minutes. Following a 10 minute centrifuge at 900xg at room temperature, the top aqueous layer is poured into a 16 mL polypropylene tube containing 6 mL of isopropanol where DNA precipitation occurs.

The precipitated DNA is immediately removed with a glass hook and transferred to a 5 mL disposable tube containing 1-2 mL of 76% ethanol and 0.2 M sodium acetate for 20 minutes. The DNA is then rinsed on the hook briefly in a microfuge tube containing 1 mL 76% ethanol and 10 mM ammonium acetate before being transferred to a microfuge tube containing 400 uL of TE buffer (10 mM Tris pH 8.0 and 1 mM EDTA) and placed on a rocker overnight at 4°C. The next day, undissolved solids is removed by centrifugation at high speed for 10 minutes. The DNA-containing supernatant is then pipetted into a new microfuge tube and stored at 4°C.

The concentration of DNA in the sample is determined by measuring absorbance at 260 nm with a spectrophotometer. Approximately 8 ug of DNA is digested with either of the restriction enzymes *Bam*H1 or *Eco*R1 as suggested by the manufacturer (Bethesda Research Laboratory, Gaithersburg, MD). This combination of enzymes cuts out the potato papain inhibitor gene intact. The DNA is then fractionated on a 0.8% agarose gel and transferred onto nylon membranes as suggested by the manufacturer (Schleicher and Schuell, Inc., Keene, NH). A potato papain inhibitor gene fragment from pDAB1400 is used as a probe. Probe DNA is prepared by random primer labeling with an Oligo Labeling Kit (Pharmacia LKB Biotechnology, Inc, Piscataway, NJ) as per the supplier's instructions with 50 microCuries 32-P-dCTP. Blots are then washed at 60°C in 0.25 x SSC (30 mM sodium chloride, 3.0 mM sodium citrate) and 0.2% sodium dodecyl sulfate for 45 minutes, blotted dry, and exposed to XXAR-5 film overnight with two intensifying screens.

To assess resistance to insect attack, transgenic plants expressing the maximal levels of midgut-effective plant cystatin are grown in 12" pots in soil. The soil is infested with *Diabrotica virgifera* eggs and the plants monitored for viability, height, root mass and standability over the course of 4 weeks. Plants expressing midgut-effective plant cystatin are significantly protected from the effects of *Diabrotica* larval damage. Alternatively, transgenic plants and populations of transgenic plants expressing PPI are assessed for *Diabrotica* resistance by the methods detailed in "Methods for the Study of Pest *Diabrotica*" (1986) eds., J. L. Krysan and T. A. Miller, Springer-Verlag, New York, pp 172-180.

Example 12: Cotransformation of Maize with PPI gene and Potato Carboxypeptidase Inhibitor

The procedures of Example 11 is substantially repeated, with the exception that a second maize tissue is transformed with a potato carboxypeptidase inhibitor (CPI) gene. The potato CPI gene encodes a protein having a sequence of Glu-Gln-His-Ala-Asp-Pro-Ile-Cys-Asn-Lys-Pro-Cys-Lys-Thr-His. Corresponding coding and non-coding oligonucleotide sequences are synthesized using a Model 380A DNA synthesizer (Applied Biosystems, Foster City, CA), with phosphoamidite chemistry. After removal of excess reagent by phenol/chloroform extraction, chloroform extraction, and ethanol precipitation, the fragments are dissolved in T4 ligase buffer. The mixture is heated to 85°C, then slowly cooled to 15°C and maintained at 15°C for at least 4 hours to allow the fragments to anneal.

pDAB303 is then digested with an excess of a restriction endonuclease recognizing a unique restriction site on pDAB303 and the larger vector fragment purified by agarose gel electrophoresis. The resulting fragment and the synthesized carboxypeptidase gene are incubated with T4 DNA ligase.

This plasmid is transformed into maize cells and stably transformed maize cells are regenerated following the procedures in Example 11.

Seed from the regenerated plants from Examples 11 and Example 12, respectively, is germinated and the resulting plants self-pollinated, i.e., pollen from one plant is used to fertilize itself to produce an S_0 population. S_1 seed from the resulting ears are grown and self-pollinations made to produce S_2 germplasm. Crosses are made between selected S_2 plants derived from Examples 11 and Example 12, respectively. The resulting F_1 hybrids are evaluated for their co-expression of PPI and CPI.

To assess resistance to insect attack, transgenic maize plants are grown and infested with *Diabrotica virgifera* eggs, as discussed in Example 11.

Example 13: Transformation of Rice Plants

A gene having the sequence set forth in Figure 1 is introduced into rice plants (*Oryza sativa*) using substantially the procedures set forth in Christou, *et al.* (1991), Bio/Technology, 9:957-962. The gene is inserted into the rice tissue culture using the transformation techniques set forth in Example 11.

To assess resistance to insect attack, transgenic rice plants are grown and infested with rice weevil, using substantially the procedures of Example 11.

Example 14: Transformation of Cotton Plants

A gene having the sequence set forth in Figure 1 is introduced into cotton (*Gossypium hirsutum* L.) plants by either particle bombardment or Agrobacterium mediated transformation.

A. Particle Bombardment

A gene having the sequence set forth in Figure 1 is introduced into cotton plants using substantially the procedures set forth in Finer and McMullen (1990), Plant Cell Reports, 8:586-589. The gene is inserted into the cotton tissue culture using the transformation techniques set forth in Example 11.

To assess resistance to insect attack, transgenic cotton plants are grown and infested with boll weevil, using substantially the procedures of Example 11.

B. *Agrobacterium*-mediated Transforamtion

A gene having the sequence set forth in Figure 1 is introduced into cotton plants using substantially the procedures set forth in Firoozabady, *et al.* (1987), Plant Molecular Biology, 10:105-116. The binary vector used for transformation, pH707-Not (set forth in Figure 8), is based on the broad host range plasmid RP4. It contains the A and B borders of an octopine-type Ti plasmid flanking a unique *Not*I cloning site, the

selectable marker 19S-nptII orf 26 and the scorabl marker 35S-gus-nos. The tetracycline-resistance gene of RP4 was replaced by erthromycin- and kanamycin-resistance genes for selection in bacteria.

To assess resistance to insect attack, transgenic cotton plants are grown and infested with boll weevil, using substantially the procedures of Example 11.

Example 15: Transformation of Alfalfa Plants

A gene having the sequence set forth in Figure 1 is introduced into alfalfa plants usng substantially the procedures set forth in D'Halluin *et al.* (1990), Crop Sci., 30:86-871. The binary vector used for transformation is pH707-Not.

To assess resistance to insect attack, transgenic alfalfa plants are grown and infested with Egyptian alfalfa weevil and alfalfa weevil, using substantially the procedures of Example 11.

Example 16: Transformation of Dry Bean Plants

A gene having the sequence set forth in Figure 1, encoding potato papain inhibitor, is introduced into dry bean plants (*Phaseolus vulgaris*) using substantially the same procedures set forth in United States Patent 5,015,580.

To assess resistance to insect attack, transgenic plants are grown and infested with Mexican bean beetle, using substantially the procedures of Example 11.

Example 17: Transformation of Potato Plants

A gene having the sequence set forth in Figure 1 is introduced into potato plants (*Solanum tuberosum* L.) is effected using substantially the same procedures set forth in Shahin and Simpson (1986), HortScience, 21(5):1199-1202. The binary vector used for transformation is pH707-Not.

To assess resistance to insect attack, transgenic potato plants are grown and infested with potato beetles, using substantially the procedures of Example 11.

Example 18: Transformation of Rapeseed Plants

A gene having the sequence set forth in Figure 1 is introduced into rapeseed (*Brassica napus* L.) is effected using substantially the same procedures set forth in Charest, *et al.* (1987), Theor. Appl Genet, (1988), 75:438-445. The binary vector used for transformation is pH707-Not.

To assess resistance to insect attack, transgenic rapeseed plants are grown and infested with flea beetles, using substantially the procedures of Example 11.

Example 19: Isolation of the Multidomain Midgut-Effective Cystatin from Tomato Leaves

Leaves from seedling tomato plants are excised and the petioles placed in moist sand. The leaves are maintained under continuous illumination for four days, after which they are gently homogenized in a Waring blendor with 0.1 M sodium phosphate pH 7.5 containing 50 µg/ml each of leupeptin, antipain, and pepstatin and 1 mM ethylenediaminetetraacetate (2 ml per gram tissue_. After straining through four layers of cheesecloth, the filtrate is centrifuged at 12,000 revolutions per minute in a Sorvall GS3 rotor and the pellet taken up in 0.1 M sodium acetate, pH 4. The solution is centrifuged to remove insoluble material and the supernatant dialyzed against 0.01 M Tris pH 7.5. The dialyzate is loaded onto a Pharmacia Mono Q HR 10/10 column and eluted with a 0-300 mM sodium chloride gradient over 40 minutes, Fractions containing peak papain-inhibitory activity are pooled, dialyzed against water and concentrated. The purified polypeptide is digested and sequence information as described in Example 1. The polypeptide is shown to be an effective inhibitor of *Diabrotica* larval midgut proteases and *Diabrotica* growth by methods described in Examples 2 and 3.

Although the invention has been described in considerable detail, with reference to certain preferred embodiments thereof, it will be understood that variations and modifications can be affected within the spirit and scope of the invention as described above and as defined in the appended claims.

WHAT IS CLAIMED IS:

1. A method of protecting a plant or a part of said plant against insect infestation by one or more insects having digestive cysteine proteases, comprising presenting to a loci wherein said insect(s) is to be controlled with an inhibitory amount of an midgut-effective plant cystatin, whereby the inhibitory amount of midgut-effective plant cystatin is ingested by the insect.

2. The method of Claim 1, wherein the midgut-effective plant cystatin is a protein having more than one cystatin domain.

3. The method of Claim 2, wherein the midgut-effective plant cystatin is potato papain inhibitor or tomato papain inhibitor.

4. The method of Claim 1, wherein the midgut-effective plant cystatin is a midgut-effective plant cystatin and a synergist capable of inactivating insect digestive enzymes which inactivate the midgut-effective plant cystatin.

5. The method of Claim 4, wherein the midgut-effective plant cystatin is a peptide of a single

cystatin domain and the synergist is carboxyp ptidase inhibitor.

6. The method of Claim 3, wherein the midgut-effective plant cystatin is a protein is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

7. The method of Claims 1 through 6, wherein the midgut-effective plant cystatin is applied to a plant species other than the plant species from which the midgut-effective plant cystatin was derived.

8. The method of Claim 1, comprising inserting into the genome of the plant a sequence coding for a midgut-effective plant cystatin with a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

9. A method according to Claim 8 further comprising the steps of:

(a) culturing cells or tissues from the plant;

(b) introducing into the cells of the cell or tissue at least one copy of a gene coding for the midgut-effective plant cystatin,

(c) regenerating resistant whole plants from the cell or tissue culture.

10. The method according to Claim 9, which comprises the further step of sexually or clonally reproducing the whole plant in such a manner that at least one copy of the sequence coding for a midgut-

effective plant cystatin with a promoter sequence active in the plant is present in the cells f progeny of the reproduction.

11. The method according to Claim 10, further comprising the steps of:

(a) selecting a fertile, insect resistant plant prepared by the method of Claim 11;

(b) sexually crossing the insect resistant plant with a plant from the insect susceptible plants from the susceptible variety;

(c) recovering reproductive material from the progeny of the cross and

(d) growing resistant plants from the reproductive material.

12. The method according to Claim 11, for imparting insect resistance to a substantially homozygous population of plants of a susceptible variety, which comprises the further steps of repetitively:

(a) backcrossing the insect resistant progeny with substantially homozygous, insect susceptible plants from the susceptible variety; and

(b) selecting for expression of both insect resistance and the other characteristics of the susceptible variety among the progeny of the backcross, until the desired percentage of the characteristics of the susceptible variety are present in the progeny along with the insect resistance.

13. The method of Claim 1, wherein the plant protected from insect infestation is maize, alfalfa, cotton, rape, dry beans, potato or rice.

14. An agricultural composition containing a carrier and an insect controlling or combating amount of at least one midgut-effective plant cystatin as an active ingredient.

15. The agricultural composition of Claim 14, wherein the midgut-effective plant cystatin is a protein having more than one cystatin domain.

16. The agricultural composition of Claim 15, wherein the midgut-effective plant cystatin is potato papain inhibitor or tomato papain inhibitor.

17. The agricultural composition of Claim 18, wherein the midgut-effective plant cystatin is a midgut-effective plant cystatin and a synergist capable of inactivating insect digestive enzymes which inactivate the midgut-effective plant cystatin.

18. The agricultural composition of Claim 17, wherein the midgut-effective plant cystatin is a peptide of a single cystatin domain and the synergist is carboxypeptidase inhibitor.

19. The agricultural composition of Claim 18, wherein the midgut-effective plant cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

20. A transgenic maize plant and its sexual progeny resistant to attack by one or more of corn root worms, maize weevils, lesser grain borer and flea

beetles, said transgenic maize plant expressing an insect controlling amount of a midgut-effective plant cystatin.

21. The transgenic maize plant and its sexual progeny of Claim 20, wherein the plant comprises a DNA sequence stably incorporated into its genome, said DNA sequence having a coding region capable of encoding a midgut-effective plant cystatin downstream of a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

22. The transgenic maize plant of Claim 21, wherein the midgut-effective plant cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

23. The transgenic maize plant of Claim 22, wherein the plant further comprises a second DNA sequence stably incorporated into its genome, said DNA sequence having a coding region capable of encoding carboxypeptidase inhibitor with a promoter sequence active in the plant to cause expression of carboxypeptidase inhibitor at levels which provide amount sufficient to inactivate insect digestive enzymes capable of inactivating the protein.

24. A transgenic rice plant and its sexual progeny resistant to attack by rice weevil, said transgenic rice plant expressing an insect controlling amount of a midgut-effective plant cystatin.

25. The transgenic rice plant and its sexual progeny of Claim 24, wherein the plant comprises a DNA

sequence stably incorporated into its genome, said DNA sequence having a coding r gion capable of encoding a midgut-effective plant cystatin with a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

26. The transgenic rice plant of Claim 25, wherein the midgut-effective plant cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

27. A transgenic potato plant and its sexual progeny resistant to attack by one or more of Colorado potato beetle and three-lined potato beetle, said transgenic potato plant expressing an insect controlling amount of a midgut-effective plant cystatin.

28. The transgenic potato plant and its sexual progeny of Claim 27, wherein the plant comprises a DNA sequence stably incorporated into its genome, said DNA sequence having a coding region capable of encoding a midgut-effective plant cystatin with a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

29. The transgenic potato plant of Claim 28, wherein the midgut-effective plant cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

30. A transgenic cotton plant and its sexual progeny resistant to attack by boll weevil, said

transgenic cotton plant expressing an insect controlling amount of a midgut-effective plant cystatin.

31. The transgenic cotton plant and its sexual progeny of Claim 30, wherein the plant comprises a DNA sequence stably incorporated into its genome, said DNA sequence having a coding region capable of encoding a midgut-effective plant cystatin with a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

32. The transgenic potato plant of Claim 31, wherein the midgut-effective plant cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

33. A transgenic alfalfa plant and its sexual progeny resistant to attack by one or more of Egyptian alalfa weevil and alfalfa weevil, said transgenic alfalfa plant expressing an insect controlling amount of a midgut-effective plant cystatin.

34. The transgenic alfalfa plant and its sexual progeny of Claim 33, wherein the plant comprises a DNA sequence stably incorporated into its genome, said DNA sequence having a coding region capable of encoding a midgut-effective plant cystatin with a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

35. The transgenic alfalfa plant of Claim 34, wher in the midgut-effective plant cystatin is a protein

having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

36. A transgenic rape plant and its sexual progeny resistant to attack by flea beetles, said transgenic rape plant expressing an insect controlling amount of a midgut-effective plant cystatin.

37. The transgenic rape plant and its sexual progeny of Claim 36, wherein the plant comprises a DNA sequence stably incorporated into its genome, said DNA sequence having a coding region capable of encoding a midgut-effective plant cystatin with a promoter sequence active in the plant to cause expression of the midgut-effective plant cystatin sequence at levels which provide an insect controlling amount of the midgut-effective plant cystatin.

38. The transgenic rape plant of Claim 37, wherein the midgut-effective plant cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

39. A substantially pure midgut-effective plant cystatin, wherein the protein is composed of more than one cystatin domain and is capable of controlling insect infestation by one or more of insects having digestive cysteine proteases.

40. The substantially pure midgut-effective plant cystatin of Claim 41, wherein the cystatin is a protein having the amino acid sequence set forth in Figure 1, or a functional derivative thereof.

41. A DNA isolate capable of encoding a protein composed of more than one cystatin domain.

42. The DNA isolate of Claim 41, wherein the DNA isolate has the nucleotide sequence coding an amino acid sequence set forth in Figure 1, or a functional derivatives thereof.

43. A biologically functional expression vehicle containing a promoter effective to promote expression of a downstream coding sequence in plant cells, a DNA coding region coding for the expression in plant cells of protein composed of more than one cystatin domain and a termination sequence effective to terminate transcription or translation of the genetic construction product in plant cells, the genetic construction effective to express in the cells of the plant insect controlling amount of the midgut-effective plant cystatin.

44. The biologically functional expression vehicle of Claim 43, wherein the DNA isolate encodes an amino acid sequence set forth in Figure 1, or a functional derivatives thereof.

45. The biologically functional expression vehicle of Claim 43, wherein the expression vehicle is pDAB219Δ-Not.

46. The biologically functional expression vehicle of Claim 44, wherein the expression vehicle is pDAB219Δ-Not.

47. A host cell transformed with a biologically functional expression vehicle of any one of Claims 43 through 47.

48. The host cell of Claim of 48, wherein the plant cell is maize, alfalfa, cotton, rape, dry beans, potato and rice.

49. The transgenic plant cell of Claim 49, wherein the DNA sequence is controlled by a promoter effective to promote expression of a downstream coding sequence in plant cell, the DNA sequence coding region coding for the expression in plant cells of a midgut-effective plant cystatin and a termination sequence effective to terminate transcription or translation of the genetic construction product in plant cells, the genetic construction effective to express in the cells of the plant insect controlling amount of the midgut-effective plant cystatin to control one or more insects having digestive cysteine proteases .

Fig.1

```
          10                  20                  30
MAIVGGLVDV  PFENKVEFDD  LARFAVQDYN

          40                  50                  60
QKNDSSLEFK  KVLNVKQQIV  AGIMYYITFE

          70                  80                  90
ATEGGNKKEY  EAKILLRKWE  DLKKVVGFKL

         100                 110                 120
VGDDSTMPGG  IVNVPNPNNT  KFQELARFAI

         130                 140                 150
QDYNKKQNAH  LEFVENLNVK  EQVVAGIMYY

         160                 170                 180
ITLAATDDAG  KKKIYKAKIW  VKEWEDFKKV

         190                 200                 210
VEFKLVGDDI  AKLGGITDVP  FPNNPEFQDL

         220                 230                 240
ARFAIQVYNK  KENVHLEFVE  NLNVKQQVVA

         250                 260                 270
GMMYYITLAA  IDAGKKKIYE  TKIWVKEWED

         280                 290                 300
FKKVVEFKLV  GDDSAKTGGI  INVPNPNSPE
```

Fig. 1 (CONT.)

```
          310                 320                 330
FQDLARFAVQ DYNNTQNAHL EFVENLNVKE

          340                 350                 360
QLVSGMMYYI TLAATDAGNK KEYEAKIWVK

          370                 380                 390
EWEDFKKVID FKLVGNDSAK KLGGFTEVPF

          400                 410                 420
PNSPEFQDLT RFAVHQYNKD QNAHLEFVEN

          430                 440                 450
LNVKKQVVAG MLYYITFAAT DGGKKKIYET

          460                 470                 480
KIWVKVWENF KKVVEFKLVG DDSAKLGGII

          490                 500                 510
NVPFPNNPEF QDLARFAVQD YNKKENAHLE

          520                 530                 540
FVENLNVKEQ LVAGMLYYIT LVAIDAGKKK

          550                 560                 570
IYEAKIWVKE WENFKKVIEF KLIGDDSAII

          580                 590                 600
GGFTDVPFPN NPEFQDLARF AVQDYNKKEN
```

Fig. 1 (CONT.)

```
          610         620         630
AHLEYVENLN VKEQLVAGMI YYITLVATDA

          640         650         660
GKKKIYEAKI WVKEWEDFKK VVEFKLVGDD

          670         680         690
SAKPGGIIIV PFPNSPEFQD LARFAVQDFN

          700         710         720
KKENGHLEFV ENLNVKEQVV AGMMYYITLA

          730         740         750
ATDARKKEIY ETKILVKEWE NFKEVQEFKL


VGDATK
```

Fig. 2

```
          10                  20                  30
           *                   *                   *
AGATG ATTTC GAAGA TTTAA GGGTA TATTG

          40                  50                  60
           *                   *                   *
TAATT TTCTT GCAAA TCTGA AAAAA TATAG

          70                  80                  90
           *                   *                   *
TTCAT GTTCT TCATC TTCTT TGTAT AATTG

         100                 110                 120
           *                   *                   *
CTACA AAAAT TTGAT ACAAA ACCAC CCAAA

         130                 140                 150
           *                   *                   *
ATCTT CTCCA AATTG CCTCA AATTT TAGAT

         160                 170                 180
           *                   *                   *
ATAAG CTCCT CTCAA CATTT CGATT CTTTT

         190                 200                 210
           *                   *                   *
GACAT ATTAC CCATT CAAAA CGGAG CTCGT

         220                 230                 240
           *                   *                   *
TCGTG GTTAG TCAAA ATTAG AACTC AAGCT
```

Fig.2 (CONT.)

```
          250         260         270
           *           *           *
GATAA ACTAA CTTTA ATGGT GGTTT CAAGT

          280         290         300
           *           *           *
TGTGC TTCAA AGTTT CAACC TTTAG CTTCG

          310         320         330
           *           *           *
ATTTT CAACC TATAG TTAGT AATTT AAAAA

          340         350         360
           *           *           *
AGAGG AAAAT AATTA CCGTC AAATA AACCT

          370         380         390
           *           *           *
CAAAA AGCTA AAAGT ATTAT ATAAA TTAGG

          400         410         420
           *           *           *
ACCAC TTAGT ATGAG GGGAC TCGAG AAGGT

          430         440         450
           *           *           *
GGGGA ATATT TACCA TTTAA TGCAT TACAT

          460         470         480
           *           *           *
GGATC ACTTG ATATG AATAT TGGAT AATTA
```

Fig. 2 (CONT.)

```
          490                 500                 510
            *                   *                   *
TAAAA ATTGA GTCTT TTTAT TTTAA TTTTT

          520                 530                 540
            *                   *                   *
TTTGG TTATT TTATT TTGAA GCTAA AATCC

          550                 560                 570
            *                   *                   *
AAATC CATCA AATCA AAATA TTTAT AACGT

          580                 590                 600
            *                   *                   *
GCTAG CCTTT TTTTT CTCTA TAAAA AGCAG

          610                 620                 630
            *                   *                   *
TTATG TTCAC CTTCT TCTCA TCACA AAAAC

          640                 650                 660
            *                   *                   *
ATTCC TTCTT CTTAA TTAAG TTATT AATTA

          670                 680                 690
            *                   *                   *
TTCGC AGTG ATG GCA ATC GTA GGA GGC CTT
             M   A   I   V   G   G   L

          700                 710                 720
            *                   *                   *
GTC GAT GTT CCA TTC GAA AAC AAA GTC GAG
 V   D>  V   P   F   E   N   K   V   E
```

FIG. 2 (CONT.)

```
                730                 740                 750
                 *                   *                   *
TTT GAT GAT CTT GCT CGT TTT GCT GTC CAA
 F   D   D   L   A   R   F>  A   V   Q

                760                 770                 780
                 *                   *                   *
GAT TAC AAT CAG AAA AAT GT AAAGA ATTAT
 D   Y   N   Q   K   N>

            790                 800                 810
             *                   *                   *
TTTTC ATTTA CTTCG ATTAC ATCTT AGCTT

            820                 830                 840
             *                   *                   *
TGTTA TGAAA AGTTA CATGT CTTAG TTAAC

            850                 860                 870
             *                   *                   *
ATAAT TGATA GTGTA AAATA TCTAC ACATC

            880                 890                 900
             *                   *                   *
ATCCG TGCAC AACAT TTAAA ATGCA TTAAT

            910                 920                 930
             *                   *                   *
GTTAC AAATA AGCAG ATGAC TCTTC GAAGA

            940                 950                 960
             *                   *                   *
ATATA TAATT AATTT TGAAA TGCTT AATTA
```

FIG. 2 (CONT.)

```
          970          980          990
            *            *            *
CTTTG TGAAT AATTA GTTGA TTCAG ATGAC

         1000         1010         1020
            *            *            *
TATTC ATATT TGTTT CATTT CAACA ACATA

         1030         1040         1050
            *            *            *
TATTT TGTAT TTCAG GAT TCT AGT TTG GAG
                     D   S   S   L   E

          1060         1070         1080
             *            *            *
TTT AAA AAG GTT TTG AAC GTG AAG CAA CAA
 F   K   K   V   L   N   V   K   Q   Q>

          1090         1100         1110
             *            *            *
ATA GTT GCT GGA ATA ATG TAC TAC ATA ACA
 I   V   A   G   I   M   Y   Y   I   T

          1120         1130         1140
             *            *            *
TTT GAG GCA ACT GAA GGT GGA AAC AAG AAA
 F   E   A   T   E>  G   G   N   K   K

          1150         1160         1170
             *            *            *
GAA TAT GAA GCC AAG ATT TTG CTG AGG AAA
 E   Y   E   A   K   I   L   L   R   K>
```

FIG. 2 (CONT.)

```
                1180                1190                1200
                   *                   *                   *
TGG GAG GAC TTG AAG AAA GTT GTA GGA TTC
 W   E   D   L   K   K   V   V   G   F

                1210                1220                1230
                   *                   *                   *
AAG CTT GTT GGT GAT GAT AGT ACA ATG CCT
  K   L   V   G   D>  D   S   T   M   P

                1240                1250                1260
                   *                   *                   *
GGG GGC ATT GTC AAT GTT CCA AAC CCA AAC
 G   G   I   V   N   V   P   N   P   N>

                1270                1280                1290
                   *                   *                   *
AAC ACC AAG TTT CAA GAA CTT GCT CGT TTT
 N   T   K   F   Q   E   L   A   R   F

                1300                1310                1320
                   *                   *                   *
GCT ATT CAG GAT TAT AAT AAA AAA CAG GTT
 A   I   Q   D   Y>  N   K   K   Q>

          1330                1340                1350
             *                   *                   *
AATTA TAATT ACTTA CTCCT CTTTT ATTTT

          1360                1370                1380
             *                   *                   *
TTCGT TAATT TCATA TTTAA ATCCC GATTT
```

FIG. 2 (CONT.)

```
          1390          1400          1410
            *             *             *
CACTA TAGTA GTACC AACAT CATAC ACAAT

          1420          1430          1440
            *             *             *
CTATT TTCCA G AAT GCT CAT TTG GAG TTT GTA
              N   A   H   L   E   F   V

          1450          1460          1470
            *             *             *
GAA AAT TTG AAT GTT AAA GAG CAA GTT GTT
 E>  N   L   N   V   K   E   Q   V   V

          1480          1490          1500
            *             *             *
GCT GGA ATC ATG TAC TAT ATA ACA CTT GCG
 A   G   I   M   Y   Y>  I   T   L   A

          1510          1520          1530
            *             *             *
GCA ACT GAT GAT GCT GGA AAG AAG AAA ATA
 A   T   D   D   A   G   K   K   K   I

          1540          1550          1560
            *             *             *
 TAC AAA GCT AAG ATT TGG GTG AAG GAA TGG
  Y>  K   A   K   I   W   V   K   E   W

          1570          1580          1590
            *             *             *
GAG GAC TTC AAG AAA GTT GTA GAA TTC AAG
E   D   F   K   K   V>  V   E   F   K
```

FIG. 2 (CONT.)

```
          1600          1610          1620
             *             *             *
CTT GTT GGT GAT GAT ATT GCA AAA CTT GGG
 L   V   G   D   D   I   A   K   L   G

       1630          1640          1650
          *             *             *
GGC ATT ACT GAT GTT CCA TTC CCA AAT AAC
 G>  I   T   D   V   P   F   P   N   N

      1660          1670          1680
         *             *             *
CCC GAG TTC CAA GAT CTT GCT CGT TTT GCT
 P   E   F   Q   D   L>  A   R   F   A

      1690          1700          1710
         *             *             *
ATT CAA GTT TAT AAT AAG AAA GAG GTTA
 I   Q   V   Y   N   K   K   E>

       1720          1730          1740
          *             *             *
ATTCA AATGG CTTAC TCCTC TTTTA TTTTT

       1750          1760          1770
          *             *             *
CGCTA GTTTC ACCTT CAAAC TTATA ATATT

       1780          1790          1800
          *             *             *
CAAAT TCTGA ATCCG CTACA GTTTC TAATA

       1810          1820          1830
          *             *             *
ATTCT TTCAT TTGAA TGAAT GCCAT ATACA
```

FIG. 2 (CONT.)

```
          1840                1850                1860
            *                   *                   *
ATCTA TTTTT CAG AAT GTT CAT TTG GAG TTT
            N   V   H   L   E   F>

          1870                1880                1890
            *                   *                   *
GTA GAA AAT TTG AAC GTT AAA CAG CAA GTT
 V   E   N   L   N   V   K   Q   Q   V

          1900                1910                1920
            *                   *                   *
GTT GCT GGA ATG ATG TAC TAT ATA ACA CTT
 V   A   G   M   M>  Y   Y   I   T   L

          1930                1940                1950
            *                   *                   *
GCG GCA ATT GAT GCT GGA AAG AAG AAA ATA
 A   A   I   D   A   G   K   K   K   I>

          1960                1970                1980
            *                   *                   *
TAT GAA ACT AAG ATT TGG GTG AAG GAA TGG
 Y   E   T   K   I   W   V   K   E   W

          1990                2000                2010
            *                   *                   *
GAG GAC TTC AAG AAA GTT GTA GAA TTC AAG
 E   D   F   K   K>  V   V   E   F   K

          2020                2030                2040
            *                   *                   *
CTT GTT GGT GAT GAT AGT GCA AAA ACT GGG
 L   V   G   D   D   S   A   K   T   G>
```

FIG. 2 (CONT.)

```
          2050          2060          2070
            *             *             *
GGC ATT ATC AAT GTT CCA AAC CCA AAC AGC
 G   I   I   N   V   P   N   P   N   S

          2080          2090          2100
            *             *             *
CCC GAG TTC CAA GAT CTT GCT CGT TTT GCT
 P   E   F   Q   D>  L   A   R   F   A

          2110          2120          2130
            *             *             *
 GTT CAG GAT TAT AAT AAT ACA CAG GTCAA
  V   Q   D   Y   N   N   T   Q>

          2140          2150          2160
            *             *             *
TTATA TATAA TGACT TACTT TTAGT TTCTT

          2170          2180          2190
            *             *             *
CTTTT TTTGT TAATT TCACA TTAAA ACCTA

          2200          2210          2220
            *             *             *
TAATA TTCAA ATCTT TAATC TATTG CAGTA

          2230          2240          2250
            *             *             *
TCTAA TAATA CTTTC ATTTG AACAA ATGTC

          2260          2270          2280
            *             *             *
ATATA CAATC TACTT TTCAG AAT GCT CAT TTG
                                        L
```

FIG. 2 (CONT.)

```
        2290          2300          2310
           *             *             *
GAG TTT GTA GAA AAT TTG AAT GTG AAA GAA
 E   F   V   E   N   L   N   V   K   E

        2320          2330          2340
           *             *             *
CAA CTT GTT TCT GGA ATG ATG TAC TAT ATA
 Q   L   V   S>  G   M   M   Y   Y   I

        2350          2360          2370
           *             *             *
ACA CTT GCG GCA ACT GAT GCC GGG AAT AAG
 T   L   A   A   T   D   A   G   N>  K

        2380          2390          2400
           *             *             *
AAA GAA TAT GAA GCC AAG ATT TGG GTG AAG
 K   E   Y   E   A   K   I   W   V   K

        2410          2420          2430
           *             *             *
GAA TGG GAG GAC TTC AAG AAA GTT ATA GAC
 E   W   E   D>  F   K   K   V   I   D

        2440          2450          2460
           *             *             *
TTC AAG CTT GTT GGT AAT GAT AGT GCG AAA
 F   K   L   V   G   N   D   S   A>  K

        2470          2480          2490
           *             *             *
AAA CTT GGG GGC TTT ACC GAA GTT CCA TTC
 K   L   G   G   F   T   E   V   P   F
```

FIG. 2 (CONT.)

```
          2500                2510.....2520
            *                   *          *
CCA AAC AGC CCC GAG TTT CAA GAT CTT ACA
 P   N   S   P>  E   F   Q   D   L   T

          2530                2540          2550
            *                   *             *
CGT TTT GCT GTT CAC CAA TAT AAT AAG GAC
 R   F   A   V   H   Q   Y   N   K>  D

        2560              2570          2580
          *                 *             *
CAG GTTAT TTATA ATGAC TTGCT CATCT
 Q>

         2590              2600          2610
           *                 *             *
TCTAT TTTTT TTTTC TAGTT AATTT CACAT

         2620              2630          2640
           *                 *             *
TCAAC CCTAT AATAT TCAAA TTCAT AATCC

         2650              2660          2670
           *                 *             *
ACTAC TGTAT CTAGT ATATA ATTCT TTCAT

          2680              2690          2700
            *                 *             *
TTGAA CGAGT GTCAT ATACA ATCAA TTTTT

          2710              2720          2730
            *                 *             *
TCAG AAT GCT CAT CTG GAG TTT GTA GAA AAT
     N   A   H   L   E   F   V   E   N
```

FIG. 2 (CONT.)

```
          2740                2750                2760
            *                   *                   *
TTG AAT GTG AAA AAA CAA GTT GTT GCT GGA
 L   N   V>  K   K   Q   V   V   A   G

          2770                2780                2790
            *                   *                   *
ATG TTG TAC TAC ATA ACA TTT GCG GCA ACA
 M   L   Y   Y   I   T   F   A>  A   T

          2800                2810                2820
            *                   *                   *
GAT GGT GGA AAG AAA AAA ATA TAT GAA ACT
 D   G   G   K   K   K   I   Y   E   T

          2830                2840                2850
            *                   *                   *
AAG ATT TGG GTT AAG GTA TGG GAG AAC TTC
 K   I   W>  V   K   V   W   E   N   F

          2860                2870                2880
            *                   *                   *
AAG AAA GTT GTT GAA TTC AAG CTT GTT GGT
 K   K   V   V   E   F   K   L>  V   G

          2890                2900                2910
            *                   *                   *
GAT GAT AGT GCA AAG CTT GGG GGC ATT ATC
 D   D   S   A   K   L   G   G   I   I

          2920                2930                2940
            *                   *                   *
AAT GTT CCA TTC CCA AAC AAC CCC GAA TTC
 N   V   P>  F   P   N   N   P   E   F
```

FIG. 2 (CONT.)

```
          2950                2960                2970
            *                   *                   *
CAA GAT CTT GCT CGT TTT GCT GTT CAA GAT
 Q   D   L   A   R   F   A   V>  Q   D

        2980                2990                3000
          *                   *                   *
TAT AAT AAG AAA GAG GTTA ATTAA AATGA
 Y   N   K   K   E>

        3010                3020                3030
          *                   *                   *
CTTAC TCCTT CTTAA TTTTT TCGTT AGTTT

        3040                3050                3060
          *                   *                   *
CACAT TCAAA TCTAT AATAT TCAAA TCCAG

        3070                3080                3090
          *                   *                   *
ATATT CCACT ACATT ATCTA ATAAT ACTTT

        3100                3110                3120
          *                   *                   *
CATTT GAACG AATTT CATAT ACAAT CTACT

          3130                3140                3150
            *                   *                   *
TTTCA G AAT GCT CAT TTG GAG TTT GTA GAA
        N   A   H   L   E   F   V   E

          3160                3170                3180
            *                   *                   *
AAT TTG AAT GTG AAA GAA CAA CTT GTT GCT
 N   L>  N   V   K   E   Q   L   V   A
```

FIG. 2 (CONT.)

```
                3190                3200                3210
                  *                   *                   *
GGA ATG TTA TAC TAC ATA ACA CTT GTG GCA
 G   M   L   Y   Y   I   T>  L   V   A

                3220                3230                3240
                  *                   *                   *
ATT GAT GCT GGA AAG AAA AAA ATA TAT GAA
 I   D   A   G   K   K   K   I   Y   E

                3250                3260                3270
                  *                   *                   *
GCT AAG ATT TGG GTT AAG GAA TGG GAG AAC
 A   K>  I   W   V   K   E   W   E   N

                3280                3290                3300
                  *                   *                   *
TTC AAG AAA GTT ATT GAA TTC AAA CTT ATT
 F   K   K   V   I   E   F>  K   L   I

                3310                3320                3330
                  *                   *                   *
GGT GAT GAT AGT GCA ATA ATT GGG GGC TTT
 G   D   D   S   A   I   I   G   G   F

                3340                3350                3360
                  *                   *                   *
ACT GAT GTT CCA TTC CCA AAC AAC CCC GAG
 T   D>  V   P   F   P   N   N   P   E

                3370                3380                3390
                  *                   *                   *
TTC CAA GAC CTT GCT CGT TTT GCT GTT CAG
 F   Q   D   L   A   R   F>  A   V   Q
```

FIG. 2 (CONT.)

```
          3400          3410          3420
             *             *             *
GAT TAT AAC AAG AAA GAG GT TAATT ATAAT
 D   Y   N   K   K   E>

        3430          3440          3450
           *             *             *
GAATT ACTCA TCTTT TATTT TATTC TCGTT

        3460          3470          3480
           *             *             *
AATTT CACAT TCAAA CCTAT AATAT TCAAA

        3490          3500          3510
           *             *             *
TCCAT AATCC ACTAC ATTAT CTAAT AATTA

        3520          3530          3540
           *             *             *
TTTCA TTTGA ATGAA CGTCA TATAC AATCT

        3550          3560          3570
           *             *             *
ATTTT TCAG AAT GCT CAC TTG GAG TAT GTA GAA
          N   A   H   L   E   Y   V   E>

      3580          3590          3600
         *             *             *
AAT TTG AAT GTG AAA GAG CAA CTT GTT
 N   L   N   V   K   E   Q   L   V

          3610          3620          3630
             *             *             *
GCT GGA ATG ATA TAC TAT ATA ACA CTT GTG
A   G   M   I   Y   Y>  I   T   L   V
```

FIG, 2 (CONT.)

```
                3640                3650                3660
                   *                   *                   *
GCA ACT GAT GCT GGA AAA AAG AAA ATA TAT
 A   T   D   A   G   K   K   K   I   Y

                3670                3680                3690
                   *                   *                   *
GAA GCT AAG ATT TGG GTG AAG GAA TGG GAG
 E>  A   K   I   W   V   K   E   W   E

                3700                3710                3720
                   *                   *                   *
GAC TTC AAA AAA GTT GTA GAA TTC AAG CTT
 D   F   K   K   V   V>  E   F   K   L

                3730                3740                3750
                   *                   *                   *
GTT GGT GAT GAT AGT GCA AAA CCT GGG GGC
 V   G   D   D   S   A   K   P   G   G

                3760                3770                3780
                   *                   *                   *
ATT ATC ATT GTT CCA TTC CCA AAC AGT CCT
 I>  I   I   V   P   F   P   N   S   P

                3790                3800                3810
                   *                   *                   *
GAG TTC CAA GAT CTT GCT CGT TTT GCT GTT
 E   F   Q   D   L   A>  R   F   A   V

                3820                3830                3840
                   *                   *                   *
CAA GAT TTT AAT AAG AAA GAG GTTA ATTCC
 Q   D   F   N   K   K   E>
```

FIG. 2 (CONT.)

```
          3850          3860          3870
            *             *             *
CTAAA TTAAA ATGAC TTAAT CTTCT TTTAT

          3880          3890          3900
            *             *             *
CTTCC GTTAG TTTCA TATTC AAATT TATAC

          3910          3920          3930
            *             *             *
 TATTT AAATC CTGAA TTTAC TGTTG TATCG

          3940          .3950         3960
            *             *             *
AATAA TTCTT TTATT TGAAC GAACG TCGTA

          3970          3980          3990
            *             *             *
TGCTC TATTT TTCAG AAT GGT CAT TTG GAG
                      N   G   H   L   E

          4000          4010          4020
            *             *             *
TTT GTA GAA AAT TTG AAT GTG AAG GAA CAA
 F>  V   E   N   L   N   V   K   E   Q

          4030          4040          4050
            *             *             *
GTT GTT GCT GGA ATG ATG TAC TAT ATA ACA
 V   V   A   G   M   M>  Y   Y   I   T
```

FIG. 2 (CONT.)

```
                4060          4070          4080
                   *             *             *
CTT GCG GCA ACT GAT GCT AGA AAG AAG GAA
 L   A   A   T   D   A   R   K   K   E

                4090          4100          4110
                   *             *             *
ATA TAT GAG ACC AAA ATT TTG GTG AAG GAA
 I>  Y   E   T   K   I   L   V   K   E

                4120          4130          4140
                   *             *             *
TGG GAG AAT TTC AAG GAA GTT CAA GAA TTC
 W   E   N   F   K   E>  V   Q   E   F

                4150          4160          4170
                   *             *             *
AAG CTT GTT GGT GAT GCT ACA AAG TGA AAT
 K   L   V   G   D   A   T   K   *>

            4180          4190          4200
               *             *             *
GAAAC TACTT TTATG TTGGT GTGAA ATAAA

            4210          4220          4230
               *             *             *
GCCAG TTGTT TGGTA TGAAG TTATT GTAAT

            4240          4250          4260
               *             *             *
GTTTG TGAGA AATAA AGCCA ATTGT TTGGC

            4270          4280          4290
               *             *             *
ATAAA ATTGC TATTA TGTTT GTGAG AAATA
```

FIG. 2 (CONT.)

```
          4300          4310          4320
            *             *             *
AAACC AGATC TATGT ATCTG AAAGT ATTTG

          4330          4340          4350
            *             *             *
ATTAT CTTAT TATAA GCGTA TTTGA ATAAT

          4360          4370          4380
            *             *             *
ATTTT GTTTT CAGAA AGCAC TCTAT AAAAA

          4390          4400          4410
            *             *             *
AATGT TATAT CACAC TTACT ATAAT ATGCA

          4420          4430          4440
            *             *             *
AATAA TTAAC TGAAC AAGTT GTTAG CAAAT

          4450          4460          4470
            *             *             *
GATAA TTAAA TGAAC AAGTT GTAAG CAAAT

          4480          4490          4500
            *             *             *
GATAG ATAAA TGAAC AAGTC GCAAG TAATT

          4510          4520          4530
            *             *             *
TATAA TTAAT TGGAC AAGTA GGAAG TCGTT
```

FIG. 2 (CONT.)

```
          4540          4550          4560
            *             *             *
TTTTT TTTAA AACAC AAATT AACAT TTCAA

          4570          4580          4590
            *             *             *
GCAAA TAATA TATAA TGAGT TTCTA AGCCA

          4600          4610          4620
            *             *             *
CTACT CATAA ACAAT CTACA TGTGA ACGTC

          4630          4640          4650
            *             *             *
AAACA CACGT ATGAT AAAAA CTCAA AGTCA

          4660          4670          4680
            *             *             *
ACCAA TATAC AAAAT AGCCA CAAGC TCATC

          4690          4700          4710
            *             *             *
TAAAC GAATT TGACG GAACT CAATA ATTTT

          4720          4730          4740
            *             *             *
ATTTA TTTAT ATTTT TTTTT TCTCC ACAAT

          4750          4760          4770
            *             *             *
GAAAT GAATT CAATG GTTCA ATAAA ATGGT

          4780          4790          4800
            *             *             *
CCATG GACCA CAAAT TTTAT TTTAT TTTTA
```

FIG. 2 (CONT.)

4810
*
AAAAA ATTAA ATGT

FIG. 3

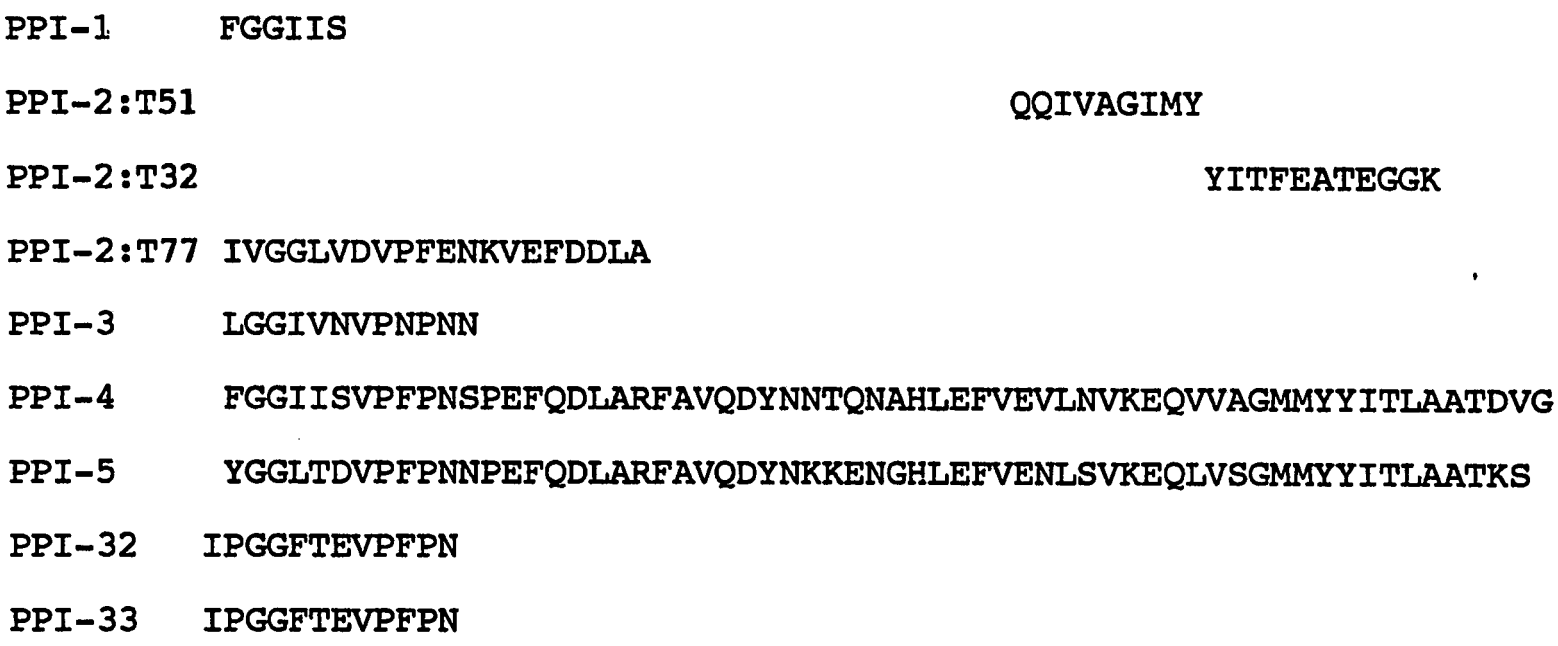

PPI-1 FGGIIS
PPI-2:T51 QQIVAGIMY
PPI-2:T32 YITFEATEGGK
PPI-2:T77 IVGGLVDVPFENKVEFDDLA
PPI-3 LGGIVNVPNPNN
PPI-4 FGGIISVPFPNSPEFQDLARFAVQDYNNTQNAHLEFVEVLNVKEQVVAGMMYYITLAATDVG
PPI-5 YGGLTDVPFPNNPEFQDLARFAVQDYNKKENGHLEFVENLSVKEQLVSGMMYYITLAATKS
PPI-32 IPGGFTEVPFPN
PPI-33 IPGGFTEVPFPN

FIG. 4

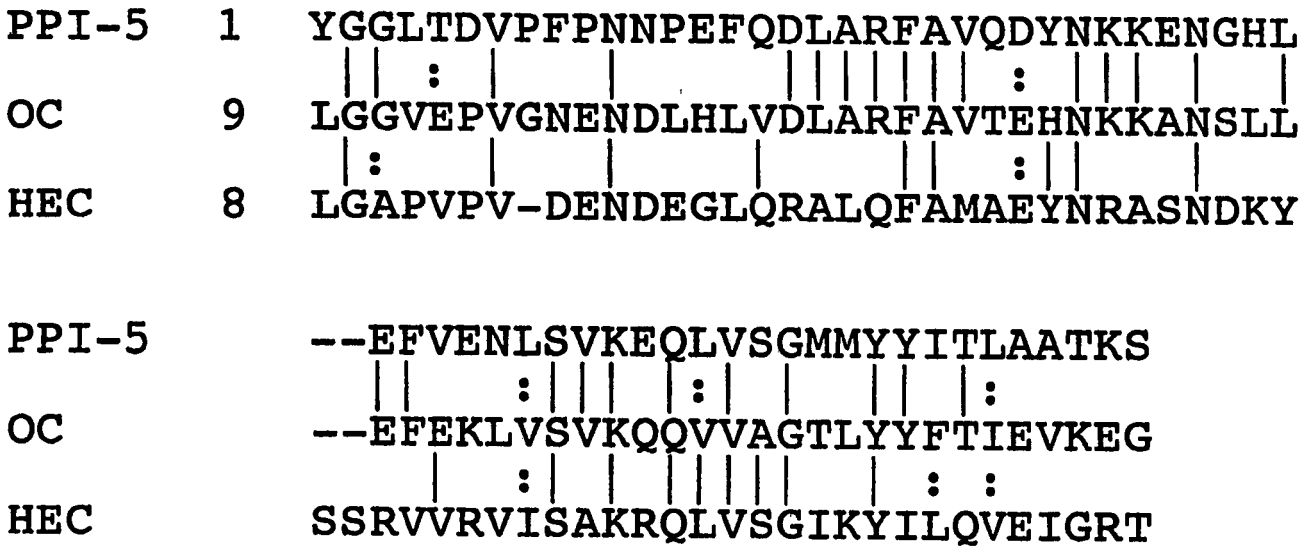

PPI-5 1 YGGLTDVPFPNNPEFQDLARFAVQDYNKKENGHL
OC 9 LGGVEPVGNENDLHLVDLARFAVTEHNKKANSLL
HEC 8 LGAPVPV-DENDEGLQRALQFAMAEYNRASNDKY
PPI-5 --EFVENLSVKEQLVSGMMYYITLAATKS
OC --EFEKLVSVKQQVVAGTLYYFTIEVKEG
HEC SSRVVRVISAKRQLVSGIKYILQVEIGRT

FIG. 5

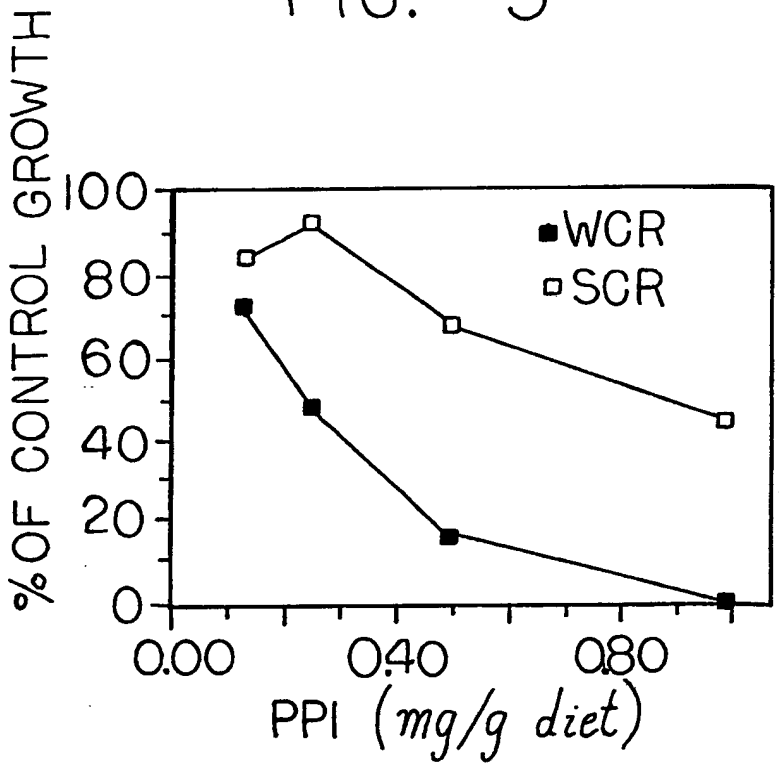

% OF CONTROL GROWTH
100
80
60
40
20
0
0.00
0.40
0.80
PPI (mg/g diet)
WCR
SCR

FIG. 6

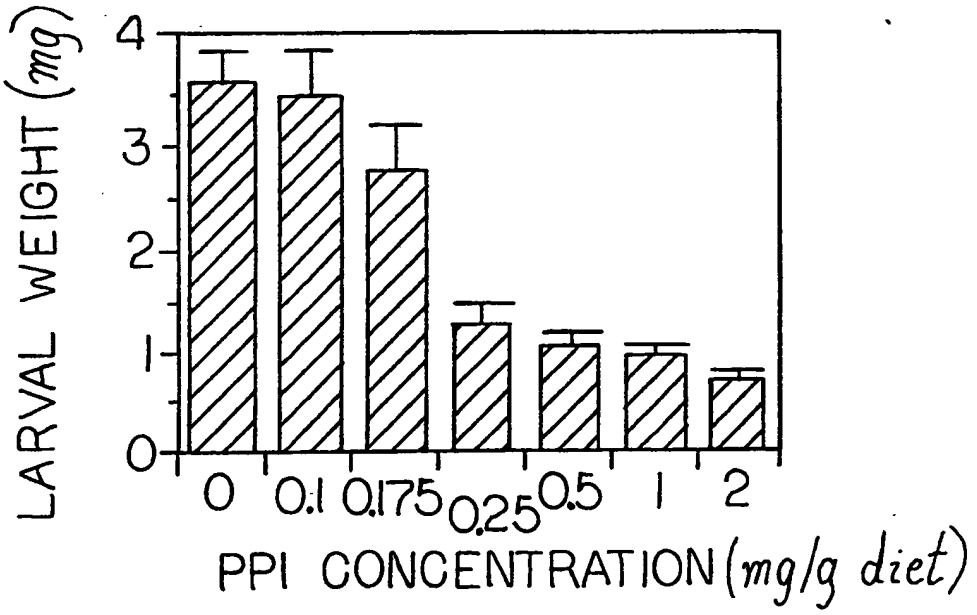

LARVAL WEIGHT (mg)
4
3
2
1
0
0
0.1
0.175
0.25
0.5
1
2
PPI CONCENTRATION (mg/g diet)

FIG. 7

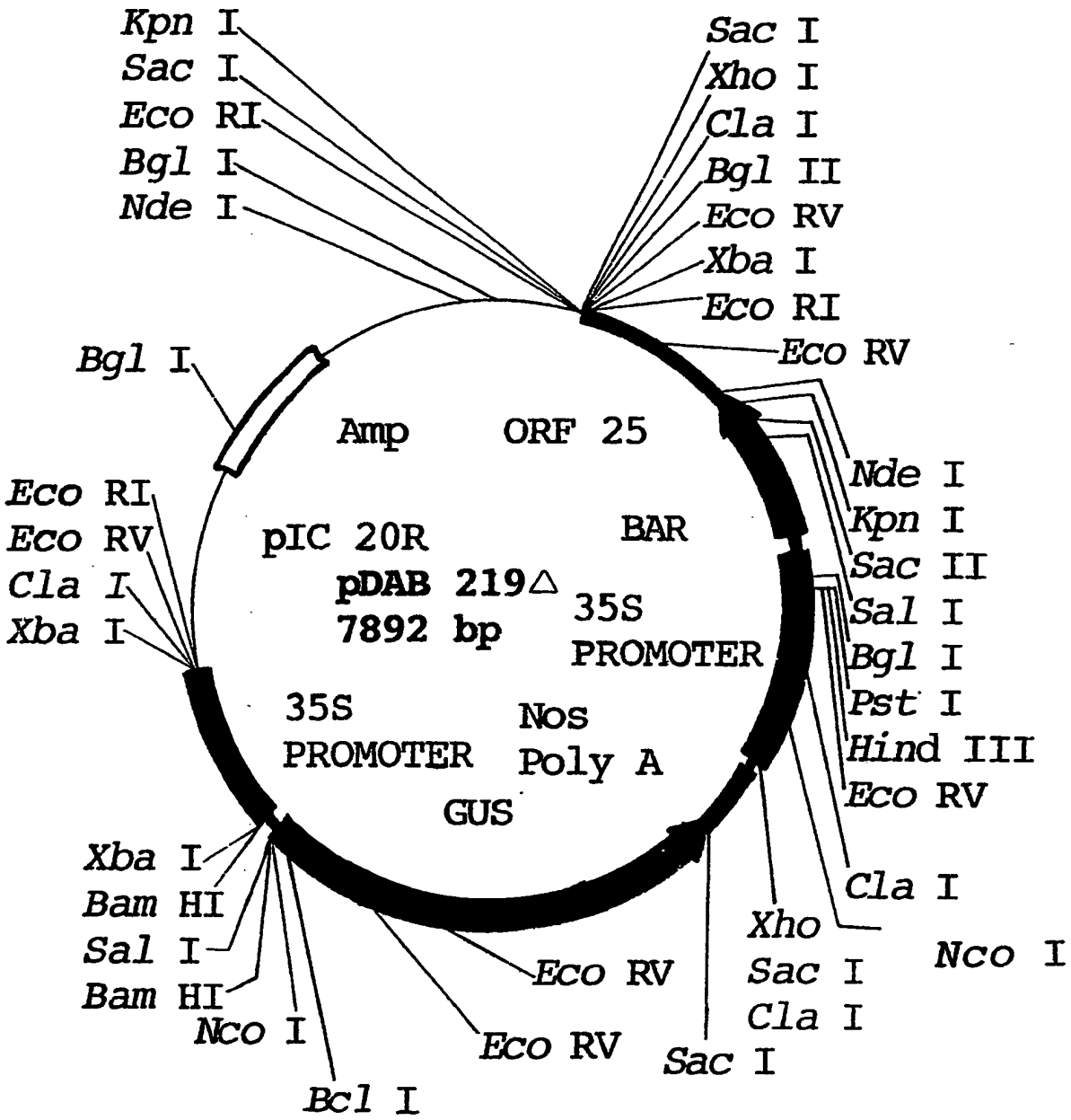
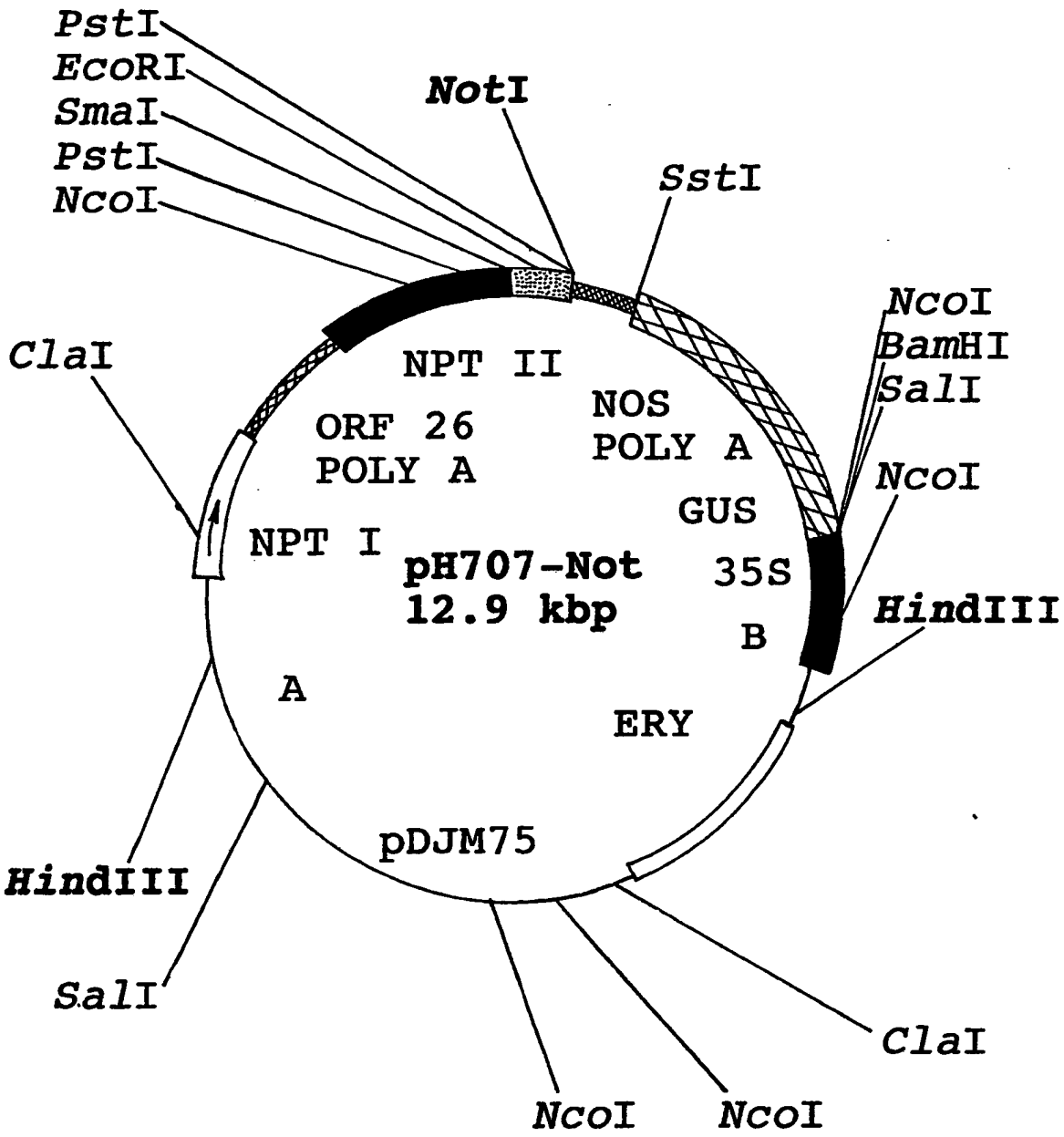

Kpn I
Sac I
Eco RI
Bgl I
Nde I
Sac I
Xho I
Cla I
Bgl II
Eco RV
Xba I
Eco RI
Eco RV
Bgl I
Amp
ORF 25
Nde I
Kpn I
Sac II
Sal I
Bgl I
Pst I
Hind III
Eco RV
Cla I
Xho
Sac I
Cla I
Nco I
Sac I
Eco RV
Eco RV
Bcl I
Nco I
Bam HI
Sal I
Bam HI
Xba I
Eco RI
Eco RV
Cla I
Xba I
pIC 20R
pDAB 219△
7892 bp
BAR
35S
PROMOTER
35S
PROMOTER
Nos
Poly A
GUS

FIG. 8


PstI
EcoRI
SmaI
PstI
NcoI
NotI
SstI
NcoI
BamHI
SalI
NcoI
HindIII
ClaI
NPT II
NOS
POLY A
ORF 26
POLY A
GUS
NPT I
pH707-Not
12.9 kbp
35S
B
A
ERY
pDJM75
HindIII
SalI
NcoI
NcoI
ClaI

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/04785

A. CLASSIFICATION OF SUBJECT MATTER

IPC(5) :C 12N 15/11, 9/50, 15/29, 1/21; A01H 1/06
US CL :435/219, 320.1, 272.3; 514/44; 800/230
According to International Patent Classification (IPC) r to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system foll wed by classification symbols)

U.S. : 435/219, 320.1, 272.3; 514/44; 800/230, DIGG56, DIGG57

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, JPOABS
SEARCH TERMS: TRANSGENIC, PLANT, PROTEASE, INHIBITOR, PAPAIN, CYSTATIN

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim N .
Y	FEBS, VOLUME 269, NO.2, ISSUED SEPTEMBER 1990, A.D. ROWAN ET AL."INHIBITION OF CYSTEINE BY A PROTEIN INHIBITOR FROM POTATO" PAGES 328-330. SEE ENTIRE DOCUMENT.	1-50
Y	TIBS, VOLUME 12, ISSUED MAY 1987, A. J. BARRETT, "THE CYSTATINS: A NEW CLASS OF PEPTIDASE INHIBITORS", PAGES 193-196, SEE ENTIRE DOCUMENT.	1-50
Y	PLANT MOLEC. BIOL., VOLUME 1, ISSUED 1983, D.E. FOERD ET AL.,"ENGINEERING OF CROP PLANTS WITH RESISTANCE TO HERBIVORES AND PATHOGENS: AN APPROACH USING PRIMARY GENE PRODUCTS" PAGES 223-233.	8-14, 21-38, 48-50
Y	J. SCI. FOOD AGRIC., VOLUME 30, ISSUED 1979, A.M.R. GATEHOUSE ET AL., "BIOCHEMICAL BASIS OF INSECT RESISTANCE IN Vigna unguiculata" PAGES 998-958, SEE ENTIRE DOCUMENT.	1-50
Y	ENTOMOL. EXP. APPL., VOLUME 41, ISSUED 1986, R.M. BROADWAY ET AL., "PLANT PROTEINASE INHIBITORS: A DEFENSE AGAINST HERBIVOROUS INSECTS?", PAGES 33-38, SEE ENTIRE DOCUMENT.	1-50

[X] Further documents are listed in the continuation of Box C. [ ] See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be part of particular relevance

"E" earlier document published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion f th international search
17 August 1992

Date f mailing of the international search report
08 SEP 1992

Nam and mailing address f the ISA/
Commissioner of Patents and Trademarks
Box PCT
Washington. D.C. 20231

Authorized officer
DAVID B. SCHMICKEL

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US92/04785

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to laim No.
Y	AGRIC BIOL. CHEM., VOLUME 52, NO.6, ISSUED 1988, M. ABE ET AL., "PURIFICATION AND CHARACTERIZATION OF A CYSTEIN PROTEINASE INHIBITOR FROM THE ENDOSPERM OF CORN", PAGES 1583-1584, SEE ENTIRE DOCUMENT.	1-50
Y	US, A, 4,940,840 (SUSLOW ET AL.) 10 JULY 1990, SEE ENTIRE DOCUMENT.	8-14,21-38, 42-50
Y	NATURE, VOLUME 328, ISSUED 02 JULY 1987, M. VAECK ET AL., "TRANSGENIC PLANTS PROTECTED FROM INSECT ATTACK", PAGES 33-37 SEE ENTIRE DOCUMENT.	8-14, 21-38, 48-50
Y	BIOTECHNOLOGY, VOLUME 7, ISSUED DECEMBER 1989, X. DELANNEY ET AL., "FIELD PERFORMANCE OF TRANSGENIC TOMATO PLANTS EXPRESSING B.T. VAR. KUSRSTAKI INSECT CONTROL PROTEIN", PAGES 1265-1269, SEE ENTIRE DOCUMENT.	1-50
Y	AU, A, 36568/89 (CIBA GIGY) 21 DECEMBER 1989, SEE ENTIRE DOCUMENT.	1-50